Published in final edited form as:

Am J Med Genet B Neuropsychiatr Genet. 2005 January 05; 132B(1): 76–84. doi:10.1002/ajmg.b.30073.

Mapping Genes of Complex Psychiatric Diseases in Daghestan Genetic Isolates

Kazima B. Bulayeva^{1,2,*}, Suzanne M. Leal³, Tatiana A. Pavlova¹, Ruslan M. Kurbanov^{2,4}, Stephen J. Glatt⁵, Oleg A. Bulayev^{1,2}, and Ming T. Tsuang^{5,6}

¹Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia

²Institute of History, Archeology, and Ethnology, Daghestan Research Center, Russian Academy of Sciences, Makhachkala, Daghestan, Russia

³Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas

⁴Daghestan State Psychiatric Hospital, Makhachkala, Daghestan, Russia

⁵Department of Psychiatry, University of California, San Diego, La Jolla, California

⁶Harvard Institute of Psychiatric Epidemiology and Genetics, Harvard Departments of Epidemiology and Psychiatry, Boston, Massachusetts

Abstract

Genetic isolates, which provide outstanding opportunities for identification of susceptibility genes for complex diseases, can be classified as primary (having an ancient demographic history in a stable environment) or secondary (having a younger demographic history) Neel [1992: Minority populations: Genetics, demography, and health, pp. 1–13]. Daghestan contains 26 out of 50 indigenous Caucasus ethnicities that have been in existence for hundreds of generations in the same highland region. The ethnic groups are subdivided into numerous primary isolates. The founder effect and gene drift in these primary isolates may have caused aggregation of specific haplotypes with limited numbers of pathogenic alleles and loci in some isolates relative to others. These are expressed as inter-population differences in lifetime prevalence and features of certain complex clinical phenotypes and in patterns of genetic linkage and linkage disequilibrium (LD). Stable highland and ethnic-cultural environments have led to increased penetrance and a reduced number of phenocopies, which typically hamper the identification of any susceptibility genes for complex diseases. Owing to these characteristics of the primary isolates, a comparative linkage study in the primary isolates allows us to define the number of susceptibility genes for any complex disease and to identify the source of variability and non-replication of linkage analysis results. As part of an ongoing study, seven extended schizophrenia and one nonspecific mental retardation kindreds have been ascertained from Daghestan isolates. Lifetime morbid risk for schizophrenia in the isolates varied from 0 to 5%. A genome scan with markers spaced 10 cM apart was carried out on these pedigrees and linkage analysis was performed using descent graph methods, as implemented in Simwalk2. To identify regions containing susceptibility genes within these kindreds, we followed up those regions with nonparametric and parametric linkage analyses,

*Correspondence to: Kazima B. Bulayeva, Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 117809 Russia. bulaeva@vigg.ru.

with the choice of genetic model guided by the results obtained in the NPL. While the analyses are ongoing, the most positive findings were made in different isolated pedigrees on chromosomes 17p11, 3q24, and 22q for schizophrenia and on chromosome 12q for nonspecific mental retardation.

Keywords

schizophrenia; mental retardation; genetic epidemiology; linkage analysis; consanguineous marriages; primary and secondary genetic isolates

INTRODUCTION

Mapping genes for complex diseases is one of the major priorities of modern genetics 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. Twin and family data support a considerable contribution of genetic factors to the predisposition to a large number of complex diseases. Genetic mapping of susceptibility genes for cardiovascular, autoimmune, and mental diseases is a rapidly developing field [Terwilliger and Ott, 1994; Peltonen, 2000; Tsuang et al., 2001; Prasad et al., 2002]. However, identification of risk genes for these complex diseases remains a complicated problem because of their genetic heterogeneity often combined with clinical heterogeneity, incomplete penetrance, and phenocopies.

Genetic isolates provide an outstanding opportunity for identification of susceptibility genes for complex diseases. Genetic isolates can be classified either as primary with ancient demographic history in a stable environment or as secondary with a younger demographic history [Neel, 1992]. The differences in demographic ages between the isolated populations are directly connected with differences in numbers of meioses and recombinations that allow using younger secondary isolates for initial detection of a disease-gene locus and older primary isolates for fine-mapping of the locus [Jorde, 2000; Bulayeva et al., 2002]. The primary isolates are located in stable geographic and ethnic-cultural environments and characterized by a limited number of ancestors, small and constant total size, endogamy, and inbreeding for hundreds of generations. A prolonged isolation in stable environment favored high genetic diversity between these primary isolates and low allelic or locus heterogeneity within them. The stable environments lead to increased penetrance and reduced numbers of phenocopies, which hamper the identification of any susceptibility genes for complex diseases. The founder effect and gene drift in the isolates caused aggregation of specific haplotypes with limited numbers of pathogenic alleles and loci in some isolates compared to others. These expressed as inter-population differences in lifetime prevalence and features of certain complex clinical phenotypes, as well as in their patterns of genetic linkage and LD. Such primary isolates still exist among ancient indigenous ethnic groups that are living in extreme and stable environments, such as deserts, highlands, jungles, or the Far North in Russia (e.g., African Saharan tribes, numerous small ethnic groups in the Caucasus highlands, or small ethnic groups in Far North Russia or Alaska, etc.), while secondary isolates are composed of historically younger migrants (e.g., Finnish isolates or Afrikaners)

or religion sects (e.g., Amish living in North America or Mormons). It is well 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 with outbred genetically heterogeneous populations. Due to these characteristics of the primary isolates, a comparative linkage and LD study of the same complex disease in such isolates may be one of most effective strategies for mapping genes of that complex disease in the presence of genetic heterogeneity. This will allow us to define the number of susceptibility genes for the disease and to identify the source of variability and non-replication of linkage analysis results.

Among the most interesting primary isolates are those living in Daghestan, located in the Northern Caucasus of Russia. Our pioneering genetic study of Daghestan ethnic populations showed the significant genetic similarity between these ethnic groups and the possibility that they are descendents of ancient ancestor population that has been fragmented into endogamous communities [Dubinin and Bulayeva, 1982; Bulayeva, 1991; Bulayeva et al., 2003a]. The results of our study of genetic diversity of Daghestan and European ethnic groups allowed us to deduce that Daghestan populations were established even earlier than European groups [Bulayeva et al., 2003a]. The Daghestan isolates meet several criteria that make them ideal for mapping genes of any complex phenotype, including:

1. This region contains 26 of 50 Caucasus small indigenous ethnic groups, most of which have been in existence hundreds of generations in stable highland ethnic and cultural environments that result in increased penetrance and reduced phenocopying, both of which should make identification of the susceptibility genes easier.
2. The population-genetic structure and demographic history of these ethnic groups have been well studied in our long-term investigation, and this information is available on the causes of inter-population differences in prevalence of certain complex diseases.
3. The prolonged reproductive isolation in severe highland environments during hundreds of generations favored high genetic diversity between these isolates and low heterogeneity within them.
4. Founder effects and gene drift favored high aggregation of certain clinical phenotypes, which should be due to a limited numbers of pathogenic alleles and loci in some isolates compared to others.
5. The ethnic tradition to keep and transmit to the next generation information on the seven generations of ancestors allows for collection of detailed genealogical data and reconstruction of extended multigenerational pedigrees for 9–12 generations.

6. Most genetic isolates are essentially one extended multi-generation kindred with a limited number of ancestors, which allows for identification of sufficient numbers of affected cases and collection of good quality clinical data.

This article describes the preliminary results of linkage studies of schizophrenia and nonspecific mental retardation (i.e., the mental retardation that was not related to gross chromosomal or known single gene mutations) in several primary and secondary genetic isolates from Daghestan.

MATERIALS AND METHODS

Description of Populations

The Caucasus region between the Black and Caspian Seas is characterized by extreme cultural and linguistic differentiation in a small geographic area that suggests a complex history. The Caucasus Mountains are astride of what must have been a major corridor of movement since the expansion of modern humans and may have functioned as a refuge and cul-de-sac off these migration streams. All these ethnic groups belong to the four linguistic families-Daghestan-Nakh, South Caucasian, Indo-European, and Altaic. The southern two-thirds of Daghestan is in the Main Caucasus Mountains, reaching 2,000–4,000 m above 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 square kilometers has a population of about two million people. While some of them are urban, there are many isolated ethnic groups that rely for subsistence on agriculture, herding, and craft production, especially in the difficult Caucasus Mountains part of Daghestan.

The total number of Daghestan's ethnic groups that were studied in our expeditions are based on the 1996 census: Avars (380,768), Dargins (321,564), Kumiks (259,302), Laks (90,146), and Tindals (3,753). Observed ethnic populations are shown on the map of Daghestan (Fig. 1). These ethnicities are subdivided into local villages that can be classified as primary isolates according to Neel [1992] definition because, in each of these villages, people have lived in the same highland environment for hundreds of generations; total population size has remained stable, and endogamy and inbreeding have been present [Bulayeva, 1991]. The total size of the isolates varies from 600 to 3,000. The smallest ethnic groups are often limited to a single unilingual village.

Clinical Characterization

In expeditions of 1995–1999, we ascertained four isolates of ethnic Dargins, one of ethnic Avars, one of ethnic Laks, one of ethnic Tindals, and one of ethnic Kumiks with aggregation of schizophrenia, as well as one isolate of Dargins with aggregation of non-specific mental retardation. Patients with schizophrenia and nonspecific mental retardation, diagnosed in Daghestan mental hospitals, were included in the study. Information from several sources was used for ascertainment of isolates with aggregation of these diseases: (1) Daghestan State Statistic Department official medical statistical reports published in 1995–1998; (2) the regional medical register of the Daghestan psychiatric hospitals that includes only data for patients diagnosed during hospitalization; (3) our own clinical observations of the affected subjects and their family members in the highland villages using a Russian translation of a

structured psychiatric interview, the Diagnostic Interview for Genetic Studies (DIGS), which was developed in the NIMH Human Genetics Initiative and uses diagnostic criteria proposed in the Diagnostic and Statistical Manual of Mental Disorders, version IV (DSM-IV); and (4) information obtained from multiple family informants on the families' history for 9–12 generations as determined by the Family Interview for Genetic Studies (FIGS), also developed for the NIMH Human Genetics Initiative.

Patients with schizophrenia and nonspecific mental retardation, diagnosed in Daghestan mental hospitals, were included in the study. Dr. Ruslan Kurbanov and Dr. Uma Guseinova, psychiatrists from the Daghestan Central Psychiatric Hospital that participate in our expeditions, interviewed all affected or probable affected subjects. The diagnoses were refined using the DIGS. The risk that schizophrenia would develop during the lifetime of a person in a local isolate was measured by the epidemiological index of lifetime morbid risk (MR) [Gottesman and Shields, 1982].

All interviews were carried out in Russian, which is the “common language” used by the Daghestan ethnic populations since the early 1900s. For the few subjects who did not speak Russian (older individuals), local medical doctors or nurses were used to carry out the interviews in the subjects' native languages. Subjects gave informed consents approved by Daghestan IRB (Daghestan Center of Russian Academy of Sciences) before participation in our study.

Population Genetic Structure

Because our approach to gene-mapping involves the aggregation of data from ethnically and geographically diverse isolated populations, it is essential to first understand the genetic diversity within and between these populations, as well as the genetic distances between these Daghestan ethnic groups and worldwide ethnic populations from Jorde' lab [Bulayeva et al., 2004]. Because of well known high mutation rate of STR loci, the genetic distances between our ethnic groups should be calculated as generalized square-sum differences between mean allele sizes in pairs of populations, with each squared difference being divided by the interpopulation variance of the sizes of alleles of individual loci [Harpending and Eller, 1999]. The size distributions of short tandem repeat polymorphisms (STRPs) can be informative about demographic history of populations. If a population is small and isolated, it loses genetic diversity relative to a large population. In the case of STRPs, there are two natural statistics describing genetic diversity: allele size variance and per locus heterozygosity. These two statistics respond to size changes at different rates: heterozygosity changes relatively rapidly while allele size variance “catches up” at a slower rate. Hence there is information about past changes in the ratio of these two statistics [Kimmel et al., 1998]. STRP allele size variance was estimated assuming a stepwise mutation model for STRPs, which appears to be consistent with the distribution of allele sizes [Di Rienzo et al., 1994; Shriver et al., 1995; Jorde, 2000]. Population comparisons of average variances across loci can be misleading, however, because the average variance tends to be dominated by a small number of highly variable, rapidly mutating STRPs [Jorde, 2000]. To control for this effect, the average variance of each locus across the three continental populations was estimated, and then a variance ratio was constructed by dividing each individual continental

variance by the average variance across continental populations. This variance ratio was then averaged across all loci for each of the three continental groups and Daghestan ethnic groups [Bulayeva et al., 2004].

Since the numbers of STR loci and alleles were large and were analyzed in several groups (ethnics, ages, age of onset groups, etc.), we used multivariate discriminate analysis to estimate the differentiation pattern of the genotypic pool of the groups studied in the multidimensional space of STRs. Discriminant function analysis is used to determine which variables discriminate between two or more naturally occurring groups. This type of multivariate analysis is usually used to determine the most informative variables that differentiate the groups studied and to distribute the subjects among these groups according to the values of these variables. Afterwards, these most informative variables (in the given case, DNA markers) may serve as predictors for group differentiation. In the case of analysis of three or more groups, $n - 1$ discriminate functions are found in the multidimensional space of the variable studied (n is the number of groups analyzed). The first function estimates the differences between the first, second, third, and the following groups; the second function, between the second, third, and the following groups, etc. All of these functions are orthogonal, and their statistical significance is estimated using the χ^2 test. The mean values of the discriminate functions may be used to determine the groups that are best identified by each of these functions in a three-dimensional space.

Linkage Analysis

A 10 cM genomic scan using Weber/CHLC 9.0 markers of 300 members of pedigrees exhibiting aggregation of schizophrenia and mental retardation from Daghestan highland isolates was performed at the Mammalian Genotyping Service of the National Institutes of Health (US) [Weber et al., 1993]. Approximately 100,000 genotypes located on 22 autosomes and X and Y-chromosomes were generated in the isolates pedigrees. In this article, we are presenting preliminary results of linkage analysis performed for several selected chromosomes in some of the genetic isolates.

Multipoint linkage analysis was performed with a statistical genetics application SimWalk2 that uses Markov chain Monte Carlo and simulated annealing algorithms that are suitable for analyzing large complex pedigrees [Sobel and Lange, 1996]. Data were analyzed using a model-free nonparametric linkage method (NPL), also known as allele-sharing statistics, to avoid false-negative findings because of misspecification of genetic parameters. Although NPL is less powerful than a correctly specified parametric linkage analysis, it is more robust when the genetic model is unknown [Risch and Giuffra, 1992]. This analysis in SimWalk2 yields statistics A, B, C, D, and E that measure the degree of clustering among the affected founder-alleles, i.e., the marker alleles descending from the founders [Sobel and Lange, 1996]. The statistic A is the number of different founder-alleles contributed to affected individuals and it is the most powerful at detecting linkage to a recessive trait. Statistic B is the maximum number of alleles among those affected, descended from any one-founder allele, and it is the most powerful at detecting linkage to a dominant trait. Statistic C is the 'entropy' of the marker alleles among those affected. Statistic D is the extent of the allele sharing among all affected pairs as measured by their identity-by-descent (IBD) kinship

coefficient. Statistic E is the NPL statistic as implemented in GeneHunter [Whittemore and Halpern, 1994].

If the NPL showed a positive signal, indicating linkage between locus of a specific chromosomal region and the clinical phenotype studied parametric analysis with two models of inheritance and penetrance (one dominant, and the other recessive) was performed for each clinical phenotype [Terwilliger and Ott, 1994; Kruglyak et al., 1996; Sobel and Lange, 1996]. Running such parametric linkage analyses, along with NPL, in different genetic isolate pedigrees for one clinical phenotype is a powerful approach toward balancing false-negative and false-positive results.

RESULTS

Figure 2 shows the genetic distances between Daghestan ethnic groups that we studied and worldwide ethnic populations, and illustrates that the Daghestan ethnic groups are closest to Europeans and less like African and Asian populations than are even some European ethnic populations. The results obtained in terms of STRP allele size variances showed that all of Daghestan's ethnic populations are prominent outliers in allele sizes variances in terms of genetic distance from African Centroid (Fig. 3).

The demographic and epidemiological description of the eight isolates included in the study, with aggregation of the mental diseases ascertained in five of Daghestan's ethnic groups, is shown in Table I. The results obtained show the lifetime morbid risk (MR) of schizophrenia in these isolates where aggregation of schizophrenia varied from 0.0143 to 0.0495 (Table I). The coefficient of inbreeding varied from 0.0017 in the Kumik population to 0.0133 in the isolate of the smallest ethnicity, Tindals.

An important clinical finding was that certain Daghestan isolates tended to have certain types of schizophrenia. For instance, the isolate no. 6022 is 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 un-appropriate affect, no sign of catatonic type of schizophrenia. The isolate no. 6007 is 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 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 20.84 ± 0.57). Mean age of onset of schizophrenia varied in different isolates from 20.6 ± 1.95 (isolate no. 6010) to 24.5 ± 2.14 (isolate no. 6030). Both the mean value and standard deviation of the age of onset were greater in the offspring of consanguineous marriages compared to the offspring of outbred marriages (21.7 ± 5.42 vs. 17.5 ± 2.4 years; $t=2.63$, $df=36$, $P=0.012$; $F=5.09$, $=0.0$). A multivariate genetic analysis demonstrated significant differences between groups of schizophrenic patients with early

and late onsets (13–19 and 26–40 years, respectively) in allele frequencies at 28 STRP loci on chromosomes 17 and 18 (χ^2 89.76, $df=68$, $P=0.040$) (Fig. 4). The results obtained showed statistically significant differences between isolated populations in the distribution of age-at-onset groups, which connected with observed clinical differences between the isolates (χ^2 18.0, $df=10$, $P=0.05$). For instance, the isolate no. 6022 had the highest rate (85%) of group-1 with early age of onset between 13–19 years of age for affected subjects; the isolate 6015 had the lowest rate (15%) for group-1; the isolate 6007 had the highest rate (50%) of the group-3 with the latest age of onset (26–40 years) and the lowest rate of this group-3 (0%) was in isolate 6022.

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

Figure 5 shows a fragment of a pedigree with schizophrenia aggregation from the isolate with lifetime MR 4.95% (Table I). 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 Levene's test showed that the proportion of schizophrenia patients was significantly higher among the offspring of consanguineous marriages ($F=9.55$, $df=1$, $P=0.003$).

The model-free NPL analysis performed in ethnically and geographically isolated Daghestan populations with schizophrenia aggregation allowed for determining a peak NPL score for loci D3S3045 and D3S2460 on 3q13.1-q13.3 and loci D17S1303 and D17S947 on 17p11.1–12 in a pedigree of isolate no. 6022 (Table II). Subsequent parametric linkage analysis of the pedigree with a recessive mode of inheritance specified and incomplete penetrance level increased the LOD score for D17S947 to 3.73 (Fig. 6). This confirmed both the genetic linkage of the clinical phenotype studied with locus D17S947 and demonstrated the adequacy of the selected model of recessive inheritance of the schizophrenia pathogenic locus with incomplete penetrance. The neighboring loci D17S2196 and D17S1294 on 17q11.2 exhibited a distinct signal arising from alleles from a common founder (per statistic B) in the schizophrenia pedigree from isolate no. 6007 (Table II).

In expedition studies of Dargin's isolate no. 6011 we ascertained a pedigree with aggregation of nonspecific mental retardation containing 15 affected cases. All these patients were diagnosed to be moderately or severely mentally retarded in the Daghestan Psychiatric Hospital. Nine affected and two probably affected pedigree members, as well as 29 of their healthy relatives, were genotyped. The NPL analysis demonstrated that the given clinical phenotype was linked with loci D12S2070 and especially with D12S395, as evidenced by the value of the B–E statistics, with NPL scores of 1.174–3.00 (Table II). Haplotype analysis in a fragment of this pedigree indicated that mentally retarded members of the pedigree had

homozygous genotypes 44 and 22 for loci D12S2070 and D12S395, except one patient (probable recombinant) with genotypes 32 and 14 (Fig. 7).

DISCUSSION

The analysis of genetic distance (Fig. 2) that encompassed the Daghestan populations and groups belonging to the three major racial groups showed three distinct clusters—African, Asian, and European (including Daghestan) populations. Allele size variance in Daghestan populations was lower than that of other European populations (Fig. 3). This shows that these ethnic populations have indeed been relatively isolated and subject to more genetic drift than many other European populations.

In our previous study, it has been shown that a constant total population size (N_t) 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 (e.g., malnutrition, lack of medical care, etc.) [Bulayeva, 1991; Bulayeva et al., 1995]. It is known that a high frequency of consanguineous marriages between the populations' members increases the rate of severe Mendelian types of hereditary diseases with early age of onset. Our population-genetic studies showed that these Mendelian diseases in Daghestan isolates were eliminated under long term strong pressure of natural selection due to the severe environment that 'cleaned up' the population's gene pool through the deaths of the affected subjects at pre-reproductive ages [Dubinin and Bulayeva, 1982; Bulayeva, 1991; Bulayeva et al., 1995]. Similar trends in isolates of Pakistan and India have been found in genetic epidemiologic studies by Bittles et al. [2002].

Yet, mutations responsible for complex diseases persist in these isolates' current generations 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 [Bulayeva et al., 1995; Jorde, 2000]. At the same time, due to cultural traditions of early marriages in Daghestan 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 complex diseases gene(s) in the isolates and consequent aggregation of certain complex diseases within them. Our epidemiological study results demonstrated that many of the isolates are characterized by prevalence of certain complex diseases such as schizophrenia, bipolar, and unipolar depression, mental retardation, cardiovascular diseases, certain types of cancer, etc., and even, within certain diseases such as schizophrenia, clinical heterogeneity of certain subtypes [Bulayeva, 1991; Bulayeva et al., 2003b].

The total population size of isolates studied (N_t) varies from 700 to 3,000; the demographical ages known from archives and historical and archeological data for the local populations vary from 600 to 5,000 years (Table I). The demographical ages are different between and within ethnic groups: for instance, among Dargin isolates, the ages vary from 700 (no. 6010) to 3,000 (no. 6030) years. The study of demographical history of isolate no.

6010 showed that it was founded approximately 700 years ago by migrants from other remote highland Dargin isolates with limited land and water resources; thus, this isolate can be classified as a secondary one.

The lifetime MR of schizophrenia in Daghestan genetic isolates varies from a complete absence (0%), through values typical of most populations of the world (about 1%), to the world maximum of about 5% (21–23). Among the isolates with aggregation of schizophrenia, MR varies from 1.6 to 4.95% (Table I).

Most isolated and inbred highland villages studied are essentially one extended multi-generation kindred with a limited number of ancestors. The traditions of the indigenous ethnic groups to keep and transmit to their offspring information on seven generations of ancestors have allowed us to reconstruct very deep multigenerational pedigrees. The results obtained demonstrated the effect of inbreeding on aggregation of schizophrenia, and our previous studies have revealed similar results for other types of complex diseases such as hypertension, depression, mental retardation, cancer, etc., [Jorde, 2000; Bulayev and Bulayeva, 2001; Bulayeva et al., 2003b]. This suggests that recessive genes in a homozygous state are involved in the formation of the complex clinical phenotypes. It is important to note that almost all the patients with a given clinical phenotype belonged to one large pedigree with a limited number of founders. A typical pedigree fragment for the isolates is shown in Figure 5. Almost all marriages in four or five generations of this pedigree were contracted between first cousins. In the most impressive case of a nuclear family of the pedigree, eight out of nine siblings were affected by paranoid schizophrenia, and their parents were also first cousins. According to the medical records we found in a regional psychiatric hospital, the paranoid form of schizophrenia also affected the father in this family.

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 gene drift that favored significant aggregation of specific haplotypes containing pathogenic 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 Daghestan isolates about between-isolate differences in aggregation of certain subtypes of schizophrenia are consistent with earlier results reported by Victor Gindilis et al. [1989]: 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.

The mean age at onset of schizophrenia in the isolates studied was about 20 years (Table I). Both the mean value and variance of the age of onset were greater in the pooled offspring of consanguineous marriages compared to the offspring of outbred marriages. The groups of schizophrenic patients with early and late ages of onset (13–19 and 26–40 years, respectively) demonstrated significant genetic diversity in the allele frequencies (Fig. 4). The observed differences in the age of onset within and between these isolates may be of crucial significance, because they indicate that the differences in gene complexes between the isolates are involved in the predisposition to the diseases with contrasting ages of onset.

Paired with the differential distribution of schizophrenia subtypes in the age-of-onset classes, this information may increase power for detecting risk genes with an effect on particular aspects of the broader schizophrenia phenotype in enriched primary isolates. It is well known that accounting for the age of onset ensured the success in mapping the genes for Alzheimer disease, breast cancer, etc. [Terwilliger and Ott, 1994; Tsuang et al., 2001; Prasad et al., 2002].

A number of promising linkage findings have emerged for schizophrenia. Some groups have reported evidence for loci predisposing to schizophrenia on chromosomes 22q11-q12, 6p24-p22, 8p22-p21, 6q, 13q14.1-q32, 5q21-q31, 10p15-p11, 1q21-q22, and 18p [Rish and Merikangas, 1996; Tsuang et al., 2001; Prasad et al., 2002; Gottesman and Gould, 2003]. The existence of numerous regions containing individual liability alleles with average LOD scores from 1.5 to 3 is a more plausible suggestion than the existence of single major genes. One interpretation of these studies is that genes of small and moderate effect may exist for schizophrenia. Numerous studies using traditional linkage analysis led to the assumption that high-penetrance mutations causing schizophrenic or nonspecific mental retardation were extremely rare or even did not exist at all [Tsuang et al., 2001; Prasad et al., 2002; Gottesman and Gould, 2003]. Genes of minor or moderate effects should be more easily detectable in such ethnically homogeneous isolated populations as those in Daghestan, which have been isolated for long periods of time [Hastbacka et al., 1994; Kruglyak et al., 1996; Jorde, 2000]. It is anticipated that the collection of larger pedigree samples from different Daghestan genetic isolates with different ethnic backgrounds, demographic ages, and founders (but originating from a common 'meta-population' many thousand years ago) will help resolve clinical and genetic heterogeneity that have contributed to the variability and non-replication of current linkage findings in schizophrenia.

The results of NPL analysis obtained in chromosome 3 loci in various isolates (Table II) were most probably related to the genetic polymorphism of the dopamine receptor D3 gene (DRD3). The 3q13 finding has much interest, because our results support other authors findings about D3 dopamine receptors that are expressed in the nucleus of the region closely related to the prefrontal neocortical areas that are the most likely targets of modern antipsychotic drugs.

Earlier we found a linkage signal for schizophrenia in region 17p11 in one of the pedigrees with preliminary diagnoses from Daghestan mental hospitals. Our subsequent clinical examination of schizophrenic patients using DIGS software allowed us to improve the diagnoses, as well as to identify several additional affected members in the pedigree. Nonparametric and parametric linkage analyses using these improved clinical diagnoses allowed for mapping the region more precisely (17p11.1-12) with a considerably higher LOD in parametric analysis (LOD=3.73) (Fig. 6). The NPL peak in the isolate no. 6022 was determined through the A statistic of SimWalk2, indicating a recessive inheritance of the pathogenic locus linked with D17S947 (Table II). The linkage analysis performed after refining clinical diagnoses in the pedigree from isolate no. 6007 also demonstrated a significant peak in the regions neighboring D17S947, at loci D17S2196 and D17S1294, according to statistics B-E (Table II). These results require further studies with a greater number of genotyped cases, but currently support that, in both isolates no. 6022 (ethnic

Dargins) and no. 6007 (ethnic Tindals) a few founder's alleles are shared among all affected pairs as measured by their IBD kinship coefficient (Table II). The pathophysiological mechanisms of the linkages found in this region of chromosome 17 may be related to the serotonin transporter gene 5-HTT polymorphism in depression and anxiety-traits that are a significant component in a pattern of schizophrenic symptoms [Heils et al., 1995; Lesch et al., 1996; Golimbet et al., 2004]. However, data reported by different researchers on the existence of this linkage or association is contradictory; and the association, if any, is weak [Prasad et al., 2002]. Our results obtained demonstrates the significant role of 5-HTT in the schizophrenia affectation in genetically and ethnically different isolates (Table II).

The differences found in recessive (isolate no. 6022) and dominant (no. 6007) inheritance of the pathogenic locus in 17q11.2-q12 may be connected with clinical heterogeneity of schizophrenia and population-specific aggregation of certain subtypes of the disease: isolate no. 6022 has aggregation of the earliest ages of onset and disorganized type of schizophrenia, while isolate no. 6007 has aggregation of cases of schizophrenia with a later age of onset and paranoid subtype.

Linkage analysis of pedigree no. 6011 with aggregation of non-specific mental retardation demonstrated that statistic B had a peak NPL at D12S395, whereas only a weak A-statistic signal was detected (Table II). The results obtained support that all affected cases in the pedigrees are sharing the founder alleles, and the pathogenic loci most likely have a dominant inheritance. The results of the parametric analysis showed that the LOD for statistic B increased to 1.78. All affected subjects in the pedigree had homozygous genotypes 44 and 22 for loci D12S2070 and D12S395, except for one probably recombinant patient with genotypes 32 and 14 (Fig. 7).

Our long-term population-genetic study of Daghestan ethnic groups showed that most likely all these ethnic groups are descendants of one ancestor 'meta-population' that existed many thousands of years ago and fragmented into endogamous communities [Bulayeva, 1991; Bulayeva et al., 2003, 2004]. Therefore the results obtained in linkage analyses of schizophrenia for the chromosome 17 region in two different ethnic isolates allow for supposition that a 30–40 cM haplotype containing a cluster of pathogenic loci emerged in this ancestor 'meta-population.' Upon subsequent fragmentation of the ancestor population into endogamous isolates, this haplotype may have been fragmented into smaller haplotypes with fewer diseases genes. As a result, an aggregation of carriers with specific haplotype, having a limited numbers of pathogenic alleles and loci, were in some isolates compared to others. Note that similar large (20–30 cM) chromosomal regions linked with schizophrenia and other mental disorders have been found in many other studies performed in outbred populations [Terwilliger and Ott, 1994; Prasad et al., 2002; Gottesman and Gould, 2003].

The comparative population study in such isolates (and their comparison with outbred populations) may allow for reliable mapping of chromosomal regions containing pathogenic loci responsible for certain clinical phenotypes. An example of successful application of this approach is the mapping of the gene for diastrophic dysplasia (DTD), a Mendelian monogenic disease, in a pedigree from an outbred population. This allowed the DTD gene to

be localized to a 2 Mb region. The subsequent analysis of LD in a Finnish isolate narrowed this region to 40 kb [Hastbacka et al., 1994].

Thus, the comparative study of different populations with aggregation of certain complex diseases is an effective way of mapping genes of complex diseases because different human populations with different gene pools, founders, and demographic history will contain different haplotypes with pathogenic loci. This may help explain the problems of non-replication of the numerous linkage and association study results performed in different human populations. This comparative population design may include, at the first stage of mapping, the search for associations with the pathogenic locus in a relatively large DNA region in historically young outbred populations. This region can be then considerably narrowed as a result of subsequent mapping in historically old primary isolates. With Daghestan's primary isolates, we have an opportunity for studying the causes of observed inter-population differences in prevalence of various clinical types related to founder effects, gene drift, and strong isolation that define the effects of susceptibility genes for numerous complex diseases. In addition, this comparative 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.

ACKNOWLEDGMENTS

We thank J. Weber (Mammalian Genotyping Service) for performing the genome-scan of Daghestan's genetic isolates subjects, and H. Coon and B. Grosser (Department of Psychiatry, University of Utah) for support of our research project in 1997–1998.

REFERENCES

- Bittles AH, Grant JC, Sullivan SG, Hussain R. 2002 Does inbreeding lead to decreased human fertility? *Ann Hum Biol* 29:111–130. [PubMed: 11874619]
- Bulayev OA, Bulayeva KB. 2001 Genetic-epidemiological study of cardiovascular diseases in Daghestan highland isolates In: Gryglewski R, Minuz P, editors. *Proceedings of the NATO Advanced Study Institute on nitric oxide: Basic research and clinical applications*. Erice, Italy: IOS Press pp 199–201.
- Bulayeva KB. 1991 Genetic basis of human psychophysiology. Gindilis VM, editor. Moscow: Nauka 218 p.
- Bulayeva KB, Isaichev SA, Pavlova TA. 1995 Population-genetic approach to the genetics of human behavior. *Biomed Sci (UK-Russian J)* 1:417–424.
- Bulayeva KB, Pavlova TA, Kurbanov RM, Bulayev OA. 2002 Mapping genes of complex diseases in Daghestan genetic isolates. *Russ J Genetics* 38:1304–1311.
- Bulayeva K, Jorde L, Ostler C, Watkins S, Bulayev O, Harpending H. 2003a Genetic and population history of Caucasus populations. *Hum Biol* 75:837–853. [PubMed: 15018034]
- Bulayeva KB, Pavlova TA, Kurbanov RM, Leal S, Bulayev OA. 2003b Genetic epidemiology study in highland isolates of Daghestan. *Russ J Genetics* 39:333–341.
- Bulayeva K, Jorde L, Ostler C, Pavlova T, Bulayev O, Harpending H. 2004 STR polymorphism in Daghestan Indigenous ethnic populations. *Russ J Genetics* 40:691–703.
- Di Rienzo A, Peterson A, Garza J, Valdes A, Slatkin M, Freimer N. 1994 Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci* 91:3166–3170. [PubMed: 8159720]
- Dubinin NP, Bulayeva KB. 1982 A study of the genetic bases of individuality in human populations. *Dokl Akad Nauk SSSR (The USSR Acad Sci Reports)* 265:334–337.

- Gindilis VM, Gainullin RG, Shmaonova LM. 1989 Genetic and demographic patterns of the prevalence of various forms of endogenous psychoses. *Russ. J Genetics* 25:432–734.
- Golimbet VE, Alfimova MV, Shchebatykh TV, Abramova LI, Kaleda VG, Rogaev EI. 2004 Serotonin transporter polymorphism and depressive related symptoms in schizophrenia. *Neuropsychiatr Genet* 126B:1–7.
- Gottesman II, Gould TD. 2003 The endophenotype concept in psychiatry: Etymology and strategic intentions. *Am J Psychiatry* 160:636–645. [PubMed: 12668349]
- Gottesman II, Shields J. 1982 *Schizophrenia: The epigenetic puzzle*. Cambridge: Cambridge University Press.
- Harpending HC, Eller E. 1999 Kato M, editor *Human diversity and its history*. Biodiversity Tokyo: Springer-Verlag pp 301–314.
- Hastbacka J, de la Chapelle A, Mahtani M, et al. 1994 The diastrophic dysplasia gene encodes a novel sulfate transporter: Positional cloning by fine-structure linkage disequilibrium mapping. *Cell* (Cambridge, MA) 78:1073–1087.
- Heils A, Teufel A, Petri S, Seemann M, Bengel D, Balling U, Riederer P, Lesch KP. 1995 Functional promoter and polyadenylation site mapping of the human serotonin (5-HT) transporter gene. *J. Neural Transm* 102:247–254.
- Jorde L. 2000 Linkage disequilibrium and the search for complex disease genes. *Genome Res* 10:1435–1444. [PubMed: 11042143]
- Kimmel M, Chakraborty R, King J, Bamshad M, Watkins S, Jorde L. 1998 Signatures of population expansion in microsatellite repeat data. *Genetics* 148:1921–1930. [PubMed: 9560405]
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. 1996 Parametric and nonparametric linkage analysis: A unified multipoint approach. *Am J Hum Genet* 58:1347–1363. [PubMed: 8651312]
- Lesch KP, Bengel D, Heils A, Sabol SZ, Greenberg BD, Petri S, Benjamin J, Muller CR, Hamer DH, Murphy DL. 1996 Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* 274:1527–1530. [PubMed: 8929413]
- Neel JV. 1992 Minority populations as genetic isolates: The interpretation of inbreeding results In: Bittles AH, Roberts DF, editors. *Minority populations: Genetics, demography and health*. London: Macmillan pp 1–13.
- Peltonen L. 2000 Positional cloning of disease genes: Advantages of genetic isolates. *Hum Hered* 50:66–75. [PubMed: 10545759]
- Prasad S, Semwal P, Deshpande S, et al. 2002 Molecular genetics of schizophrenia: Past, present, and future. *J Biosci* 27:35–52. [PubMed: 11927776]
- Risch N, Giuffra L. 1992 Model misspecification and multipoint linkage analysis. *J Hum Hered* 42:77–92.
- Rish N, Merikangas K. 1996 The future of genetic studies of complex human diseases. *Science* 273:1516–1517. [PubMed: 8801636]
- Shriver M, Smith M, Jin L, Marcini A, Akey J, et al. 1995 A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol Biol Evol* 12:914–920. [PubMed: 7476137]
- Sobel E, Lange K. 1996 Descent graphs in pedigree analysis: Applications to haplotyping, location scores, and marker sharing statistics. *Am J Hum Genet* 58:1323–1337. [PubMed: 8651310]
- Terwilliger JD, Ott J. 1994 *Handbook of human genetic linkage*. Baltimore: Johns Hopkins Univ Press.
- Tsuang M, Stone W, Faraone S. 2001 Genes, environment and schizophrenia. *Br J Psychiatry Suppl* 40:18–24.
- Weber JL, Wang Z, Hansen K, et al. 1993 Evidence for human meiotic recombination: Interference obtained through construction of a short tandem repeat-polymorphism linkage map of chromosome 19. *Am J Hum Genet* 53:1079–1095. [PubMed: 8213834]
- Whittemore AS, Halpern J. 1994 A class of tests for linkage using affected pedigree members. *J Biometrics* 50:118–127.

Fig. 1.
Map of Daghestan showing the isolated populations (black points) studied in our expeditions.

Fig. 2.
Genetic distances between Daghestan and worldwide ethnic populations. Daghestan ethnics are framed.

Fig. 3.
Allele size variances plotted against genetic distance from the African Centroid in Daghestan and worldwide ethnic populations. Daghestan ethnics are shown on italic.

Fig. 4. Multivariate genetic analysis of schizophrenia patients grouped according to their age of onset in the space of first two discriminant functions (I and II) for 28 STR loci of Chrs 17 and 18. Age of onset groups: (1) 13–19; (2) 20–25, and (3) 26–40.

Fig. 5.

The sections of Daghestan primary isolate pedigree with aggregation of schizophrenia.

Affected subjects marked by color (solid) and probably affected by points. The double lines indicate consanguineous marriages.

Fig. 6. The multipoint Lod score for linkage of schizophrenia across chromosome 17 in pedigree from isolate 6022 (results of parametric analysis of Simwalk 2.86 using recessive model with penetrance 90%).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table I

Description of Daghestan Genetic Isolates With Aggregation of Mental Diseases

Isolates	Ethnic groups	ISLT	F	Nt	#PM	Naf(NO)	NOtot	MR	Mean AGEON	AFST
1.6005	Laks	5,000	0.0114	931	274	25 (19)	60	0.0241	24.1±2.15	SCZ
2.6007	Tmdals	3,000	0.0133	1,800	539	36 (26)	86	0.0176	24.0±2.35	SCZ
3.6029	Kumiks	600	0.0017	2,000	342	24 (18)	43	0.0163	20.7±1.61	SCZ
4.6010	Dargins	700	0.0066	2,000	533	70 (24)	50	0.0297	20.6±1.95	SCZ
5.6022	Dargins	4,000	0.0091	1,340	314	20 (16)	46	0.0271	20.8±1.51	SCZ
6.6030	Dargins	3,000	0.0110	700	257	28(9)	14	0.0495	24.5±2.14	SCZ
7.6015	Avars	800	0.0099	1,153	396	23 (21)	43	0.0143	21.7±2.05	SCZ
8.6011	Dargins	700	0.0066	2,000	187	15(11)	26	—	—	MRet

Note: Nt, the total number of the isolates; #PM, the number of pedigree members; ISLT, approximate isolation time of the populations known from archives, ancient publications, and archeological data; Naf (NO), number of affected in the pedigree (NO, number of observed among them); NOtot, number of totally observed in the pedigree for clinical (DIGS) and DNA sampling; MR, lifetime morbid risk; AGEON, mean age of onset of schizophrenia; AFST, affection status; SCZ, schizophrenia; MRet, mental retardation.

Table II

. Model Free Non-Parametric Linkage Analysis (Simwalk2) for Chromosomes 3, 12, and 17 for Schizophrenia (SCZ) and Unspecific Mental Retardation (MRet) Affection in Some of Daghestan Genetic Isolates Pedigrees

Statistics								
Chromosomes	Markers	A	B	C	D	E	Dx	#KIND
3	D3S3045	1.42 (0.038)	1.244 (0.057)	1.259(0.055)	1.268 (0.054)	1.495 (0.032)	SCZ (8/9)	6022
	D3S2460	1.41 (0.039)	0.629 (0.235)	0.836(0.146)	0.772 (0.169)	0.879 (0.132)		
12	D12S2070	1.134 (0.073)	1.174(0.057)	1.635(0.023)	1.549 (0.028)	1.406 (0.039)	MR (11/29)	6011
	D12S395	1.154 (0.070)	2.125(0.007)	2.000(0.010)	3.000 (0.001)	2.699 (0.002)		
17	D17S1303	1.186 (0.065)	0.561 (0.27)	1.414(0.036)	1.032 (0.093)	0.851 (0.141)	SCZ (8/9)	6022
	D17S947	1.575 (0.026)	0.476 (0.334)	1.836(0.014)	1.316 (0.046)	0.976 (0.105)		
	D17S2196	0.419 (0.380)	1.146(0.075)	1.046(0.090)	1.262 (0.054)	1.442 (0.036)	SCZ (11/40)	6007
	D17S1294	0.426 (0.375)	1.152(0.070)	1.061 (0.089)	1.265 (0.054)	1.282 (0.052)		

Note: Dx, clinical phenotypes with numbers of genotyped patients and healthy kindred members; #KIND, number of kindred.