ISSN 1022-7954, Russian Journal of Genetics, 2016, Vol. 52, No. 7, pp. 746–756. © Pleiades Publishing, Inc., 2016.
Original Russian Text © V.A. Stepanov, 2016, published in Genetika, 2016, Vol. 52, No. 7, pp. 852–864.

Evolution of Genetic Diversity and Human Diseases

V. A. Stepanov[a, b]

[a] *Research Institute of Medical Genetics, Russian Academy of Sciences, Tomsk, 634050 Russia*
[b] *National Research Tomsk State University, Tomsk, 634050 Russia*
e-mail: vadim.stepanov@medgenetics.ru

Received February 24, 2016

Abstract—The problem of development and dispersion of complex diseases in human populations requires new views, approaches, hypotheses, and paradigms. Evolutionary medicine provides one of the promising approaches to this problem, putting the disease into an evolutionary context. Unlike classic approaches oriented to proximate issues on structure and mechanisms of a disease, evolutionary considerations are broader. It provides the basis for understanding the origin, dispersion, and maintenance of the high frequencies of pathological phenotypes in modern human populations. In the current paper, we try to review the modern concepts on the evolution of human genetic diversity, to shape the outlines of evolutionary medicine, and to illustrate evolutionary medical problems using our experimental data. Data on genome-wide search for the signals of decanalization and adaptation in the human genome and on related biological processes and diseases are presented. Some hypotheses and concepts of evolutionary medicine may be productive for revealing the mechanisms of origin and dispersion of complex diseases and for pathogenetics of multifactorial diseases. One of such concepts is the hypothesis of decanalization of genome–phenome relationships under natural selection during modern human dispersion. Probably, the high frequency of alleles associated with complex diseases (and partially the high prevalence of diseases themselves) could be explained in the framework of the hypothesis.

Keywords: human population, genetic diversity, natural selection, adaptive evolution, multifactorial diseases
DOI: 10.1134/S1022795416070103

INTRODUCTION

Progress in describing genetic diversity of modern humans, associated with the introduction of genome-wide genotyping and sequencing technologies, has led to describing and cataloging tens of millions of polymorphic variants of the genome [1, 2]. On the other hand, the ideology of the search for the genetic component of common human diseases based on association studies of polymorphic variants of the genome, including genome-wide association studies (GWAS) and sequencing of exomes and genomes, has almost reached the limit of detection of the possible inherited component of multifactorial diseases (MDs). In the majority of MDs, the proportion of heritability that is explained by the known genetic factors identified in family, twin, associative, and other genetic studies does not exceed 10–15%. The remaining genetic variability, which underlies diseases, is the so-called "lost heritability" [3–5]. The problem of "lost heritability" hardly can be solved in the framework of the existing paradigm, despite the rapid technological advances in genomics and its derived "omics." The synthesis and crystallization of genomic data within the framework of bioinformatics approaches and systems biology is a possible solution to the problem, although examples of successful system synthesis for common diseases are scanty [6–8]. Another search reserve is to study the pathogenetics of a disease at the ontogenetic level, to study the "disease ontogeny" (epigenetic modifications of genetic information, transgenerational effects, somatic mosaicism, and paradominant inheritance mechanisms) [9].

Undoubtedly, the problem of understanding the mechanisms of emergence and distribution of MDs in modern populations requires a new view, new approaches, and new hypotheses and paradigms which do not abolish the huge strata of previous achievements but complement and extend them. One of such promising approaches in understanding the nature of human diseases is evolutionary medicine—the study of human diseases in the evolutionary context. Unlike the classical approaches, which focus on the applied problems of the structure and mechanisms of disease development, the evolutionary approach is broader and makes it possible to study the origin, distribution, and maintenance of the high frequency of pathological phenotypes in populations. In this article, we have attempted to briefly summarize the current ideas about the evolution of human genetic diversity, to outline the scope of evolutionary medicine, and to illustrate the evolutionary and medical prob-

lems by a number of modern concepts and our own data.

GENETICS OF THE ORIGIN AND DISPERSAL OF MODERN HUMANS

Current studies of the genetic diversity in human populations on the basis of molecular genetic data have made it possible to reconstruct the origin and dispersal of *Homo sapiens* on the globe. According to the generally accepted notions of the phylogeography and phylogeny of the uniparental regions of the human genome (mitochondrial DNA and nonrecombinant part of Y chromosome), the genome-wide data for single nucleotide polymorphisms (SNPs), and whole genome sequencing data, the genetic diversity of our species goes to an ancestral African population, which long inhabited the tropical climate zone and had a small abundance (from several hundred to several thousand people) [10, 11]. The convergence time of human genetic diversity by uniparental lines to the least ancient common ancestor is estimated at 100000–200000 years, and the start time of the modern human dispersal from Africa and the expansion of abundance is estimated at 50000–100000 years. Recent data on the whole genome sequencing of Neanderthal and Denisovan man revealed hybridization of *Homo sapiens* with other species of the genus *Homo* existing at the same time [12–14], thus raising doubts about the hypothesis of a recent African origin of modern humans in favor of the multiregional evolution model (hybridization model) (see [15]). However, the fixed contribution of other species to the gene pool of the modern humans is small (no more than fractions of a percent). The fact of hybridization of *Homo sapiens* with Neanderthal and Denisovan man, which can be considered proven, does not negate the idea of the recent African roots, obtained within the framework of the uniregional hypothesis, but complements it, making the picture of the evolution of our species more intricate. Notably, the modern molecular genetic data form the notions of the origin and dispersal of humans that do not contradict the data on the classical protein markers and paleoanthropological data.

ENVIRONMENTAL CHANGES AND PREREQUISITES FOR ADAPTIVE EVOLUTION OF GENETIC DIVERSITY

In the course of dispersal from Africa, modern humans, who populated all continents and all climatic zones, faced global environmental challenges, which created prerequisites for adaptive evolution of the genetic diversity [16]. The process of population of other continents and other climatic zones, which began approximately 70000 years ago and ended by the population of America approximately 20000 years ago and last migrations to remote islands of Oceania approximately 2000 years ago, required the adaptation to climate and geographical conditions of the new environment. The emergence of agriculture and the transition from the cultures of hunters and gatherers to the cultures of farmers and herdsmen, which entailed a rapid expansion of the population abundance approximately 10000 years ago, were accompanied by an abrupt change in the diet and necessitated the adaptation to the changed diet and metabolism. In the era of great geographical discoveries and migration of peoples (500 years ago), massive migrations literally within one generation dramatically changed the environment of large population groups (for example, during the colonization of America and export of slaves from West Africa to the New World). The last global challenge was the industrial revolution, which caused environmental pollution and the emergence of numerous new chemicals that previously did not exist in nature. The gene pool, as a relatively slowly evolving system, and especially the human genome are unable to change under the pressure of environmental factors for a small number of generations and quickly respond to environmental challenges. However, the traces of adaptive evolution associated with the early events of human dispersal and with the adaptation to changes in the diet in the era of the agricultural revolution may be the subject of research and genetic analysis. The downside of adaptive changes may be the "excision" of genetic markers that determine the susceptibility to common diseases, which manifests itself as a high incidence of certain multifactorial diseases.

THE STRUCTURE OF GENETIC DIVERSITY IN MODERN HUMAN POPULATIONS

The evolutionary history of modern human populations, formed as a result of human dispersal from Africa and subsequent migration events, has led to the formation of the modern characteristic patterns of genetic diversity in human populations. The most general pattern of the organization of genetic diversity in human populations is its strict geographic structuring, which manifests itself as clustering of geographically close populations in respective statistical procedures as well as the clinal changes in allele frequencies and heterozygosity level. In our opinion, the main driving forces of microevolution during human dispersal were, apparently, migrations and genetic drift together with possible local effects of selective factors [10]. The scenario of gradual population of continents assumes successive events of migrations, genetic drift, and possible isolation of populations by distance, with subsequent moderate gene flow between close populations. The role of adaptive evolution, mediated by natural selection on a genome-wide scale (or the gene pool of populations), is often hidden by other factors of population dynamics.

The geographical structuring of genetic diversity is detected in almost all representative samples of genetic

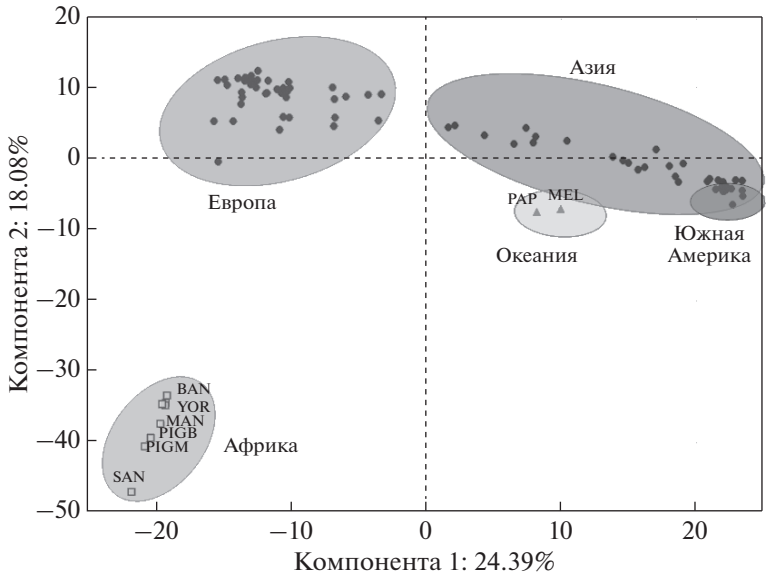

Fig. 1. The space of the principal components of allele frequencies of 200000 SNPs in 79 world populations. @Key: Компонента –> Component; Европа –> Europe; Африка –> Africa; Азия –> Asia; Океания –> Oceania; Южная Америка –> South America.

markers of any type (including the global and subregional scales) for both populations and individual genomes. An example of the geographical pattern of diversity is the analysis of 200000 SNPs performed by us in 79 world populations (unpublished data). The location of populations in the space of two first principal components of frequencies of 200000 genome-wide markers is almost identical to the geographical map on the scale of both the world (Fig. 1) and Eurasia. Individual genomes repeat the pattern that is roughly observed at several dozen SNPs randomly selected from 200000 SNPs. The two main factors that, in this case, account for more than 40% of the variability of allele frequencies in populations are the geographic parameters—latitude and longitude.

The specified distribution pattern of genetic diversity in human populations has two principal consequences, which were confirmed by actual data. Firstly, humans have a relatively low level of genetic differentiation of populations at a high level of intrapopulation diversity. The proportion of interpopulation genetic differences in the overall diversity is not more than 10–15% of the worldwide population (see [10]). In particular, in the above example of the analysis of 79 world populations for 200000 SNPs, the global level of population subdivision is 11.4% ($F_{st} = 0.114$). Secondly, at the genome-wide level or in the samples of neutral markers, a gradual (clinal) decrease in genetic diversity from the center of origin (Africa) to the periphery of the range is observed. An example of such picture for Eurasia obtained by us by means of 200000 SNPs is shown in Fig. 2.

HUMAN DISEASES IN THE EVOLUTIONARY CONTEXT

Of course, it should be expected that the evolutionarily formed pattern of genetic diversity is reflected in the part of the genetic variation that is relevant to human diseases. It can be assumed that, if the "disease genes" are not associated with adaptation to global environmental challenges, they should behave in the same way as the main part of the gene pool (i.e., geographically close populations should have a similar spectrum of mutations in monogenic diseases or a similar spectrum of the composition and frequency of the genetic markers of susceptibility to common multifactorial diseases). At the same time, if the genetic component of diseases had adaptive significance, the disturbance of the general consistent pattern of the geographical structuring of genetic diversity due to the greater role of natural selection for this part of the genome as compared to the neutral markers or the genome-wide picture can be expected for the "disease genes."

The literature provides many examples of the first pattern, for example, the similarity of the spectra and frequencies of mutations in cystic fibrosis in the populations of the European origin [17] or the ethnoterritorial structuring of the spectrum of frequencies of MD markers associated with diseases according to the genome-wide association study (GWAS) data [10, 18]. The second scenario is also identified in the experimental data both for monogenic diseases (the classic example is the wide spread of the heterozygous carriage of mutations in the hemoglobin genes in sickle cell anemia and beta-thalassemia in the areas with

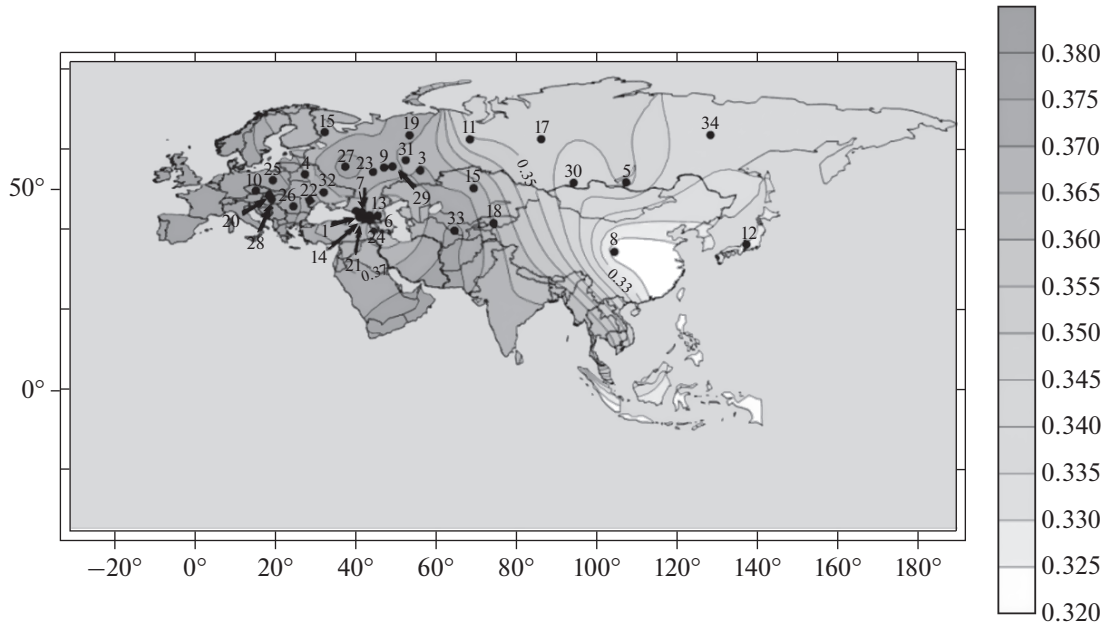

Fig. 2. Clinal change in the genetic diversity (mean heterozygosity for 200000 SNPs) in Eurasia. The numerals on the map designate the number of populations in which the mean expected heterozygosity was estimated. The scale of expected heterozygosity values is shown on the right.

malaria [11, 19]) and for MDs. An example of disturbance of the geographical structuring of marker frequencies in MDs is the frequencies of the genes that determine the susceptibility to autoimmune and immune diseases, which show the effect of decanalization of the genome–phenome relationships: the immune response is mediated by the adaptation to the global environmental factors [20].

The analysis of human diseases in the evolutionary context is of considerable interest from the standpoint of searching for genes and disease markers on the basis of evolutionary concepts and is a reserve for detecting the "lost heritability" in MDs. In this case, taking into account the evolutionary factors adds another dimension to the phenomenon of complexly inherited diseases and requires special attention in forming the working hypothesis and planning the design of the study. In the case–control group design, a group of patients with a certain phenotype is formed, which, owing to the major role of environmental factors as well as clinical and genetic heterogeneity of almost any complex disease, inevitably includes individuals with a similar phenotype but different etiology of the disease. Then, on the basis of the working hypothesis, the genetic markers are selected whose genotyping in patients and in control subjects indicates the relationship (association) between the marker and the disease. The clinical phenotypes underlying complex diseases are based on a great variety of interacting genetic and environmental factors. The genetic factors range from the alleles with minimum effect to the "major genes," probable for some phenotypes, and the environmental factors range from the population-specific to family and individual ones. The gene–gene interaction, the interactions of genes and the environment, and the epigenetic regulatory mechanisms, in turn, are also influenced by genetic control and environmental signals and are also factors that complement the overall picture of the formation of clinical phenotypes that lead to MDs. The nondirectional effects of microevolutionary factors (mutations and drift) provide random variations in the genetic components of the disease, whereas the directional effects of the environment on the genetic component of the phenotype is mediated by the action of natural selection. The attempts to identify the role of selection in the occurrence of diseases against the background of neutral evolution of genes that formed the main vector of genetic variation in human populations are among the main problems of evolutionary medicine.

Evolutionary medicine considers the health/disease relation in the evolutionary perspective. This is an attempt to understand the mechanisms of emergence and development of diseases and the relationship between genetic diversity and diseases within the framework of modern evolutionary theory and concepts of evolutionary biology. Evolutionary medicine is a conceptual approach to the analysis and comprehension of information generated by modern technologies—whole genome sequencing, transcriptomics, epigenomics, etc. The progress and gradual maturation of evolutionary medicine to a separate direction was facilitated by the achievements of modern genomics, derived "omics," and systems biology. It is the

necessity to somehow understand and explain the accumulating data from the standpoint of the health/disease relation that caused the resurgence of interest in the evolutionary medicine. Recent achievements in this direction include the first genome-wide search for signals of natural selection in the human genome [21], the discovery of human orphan genes (genes present only in one taxon (up to species)) [22], and the genome-wide study of the climate adaptation signals [23].

SOME CONCEPTS ON THE ROLE OF EVOLUTIONARY FACTORS IN HUMAN DISEASES

With regard to the microevolutionary processes in human populations, various theoretical concepts and hypotheses have been and continue to be developed which reveal the possible mechanisms of action of selection for the structure of the gene pool and phene pool of human populations, including monogenic and multifactorial diseases. Among the important theoretical concepts, the canalization concept by K. Waddington [24], the notions of the stabilizing selection by I.I. Schmalhausen [25, 26], and the concept on the optimal level of heterozygosity in population by Yu.P. Altukhov [27, 28] should be mentioned. Regarding the role of selection in the epidemiology of common diseases, the concepts of "thrifty genes" [29, 30], "drifting genes," and several others were developed. The appearance in recent years of new data based on genome-wide and postgenomic technologies has led to a resurgence in conceptual understanding of the role of natural selection in the approaches of evolutionary medicine and the development of new ideas to the search for selection signals in human populations. Among the new theoretical ideas, the model of "ancestral susceptibility" by Di Renzo and Hudson [31] and the decanalization concept by Gibson [32, 33] can be mentioned. Let us try to briefly consider these ideas and their "sphere of competence."

The canalization concept, which was developed by Conrad Waddington, the author of the term "epigenetics," in the 1940s [24] in relation to ontogenetic processes and further developed by his followers [34, 35], postulates that a system evolves to the most stable state, which is resistant to genetic and environmental changes and is represented by the optimum phenotype in the major part of the population. Such a population, which is in a canalized state, is maximally adapted to the given environmental landscape through respective epigenetic and environmental signals. The idea of the stabilizing selection by Schmalhausen [25, 26] and the more recent concept of the optimal level of heterozygosity in a population, which is autoregulated by the population dynamics factors, by Altukhov [27, 28] are close to Waddington's ideas.

Currently, Waddington's ideas are rethought owing to accumulation of large arrays of genetic and genomic data and owing to the problems of searching for the genetic basis of common diseases. In particular, the Australian researcher Greg Gibson proposed the decanalization concept to explain the evolutionary mechanisms of accumulation of common diseases in populations [32, 33]. The decanalization concept implies that human dispersal beyond the range of the ancestral populations has disturbed the channeled optimum and has led to the appearance of cryptic (hidden, not manifested in a canalized population) genetic variability and increased frequency of associated diseases. Environmental and cultural changes over the past 100 years (diet, smoking, and pollution) have even more significantly shifted physiology from the stable channeled mechanisms.

The hypothesis of "ancestral susceptibility" [31], in accordance with the canalization principle, postulates that the ancestral alleles were adaptive for ancient human populations, whereas the derived alleles were "harmful." With the dispersal and changes in the environmental factors, the pressure of selection for the derived alleles decreased, and the ancestral alleles became the alleles of susceptibility to chronic diseases. In fact, the hypothesis of "ancestral susceptibility" is a special case of "decanalization" acting against the ancestral alleles.

The concept of "thrifty genotypes" [29, 30], according to which the selection in ancient populations was aimed at accumulating the alleles that facilitate energy storage in the form of fats, which in today's world has led to a high rate of diabetes and obesity, in turn, is a special case of "ancestral predisposition."

CANALIZATION AND DECANALIZATION OF GENOME–PHENOME RELATIONS DURING THE MODERN HUMAN DISPERSAL: EXAMPLE OF GENES FOR AUTOIMMUNE AND IMMUNE DISEASES

The evolutionary precursors of *Homo sapiens* and the ancestral populations of modern humans for many generations have been subjected to increased pressure of environmental factors associated with infectious tropical diseases (helminths and bacterial infections). The role of parasite (especially helminth) infections in population morbidity is illustrated by epidemiological data. According to a number of studies [36, 37], helminthiases are much more common in the populations inhabiting the zones with hot and humid tropical climate than in the populations of the temperate and arctic zones. In tropical populations, natural selection has favored the accumulation of the gene variants that ensure a stronger immune response to parasitic infections by the Th2 type [38, 39]. This should have led to the canalization of the genetic structure of the population owing to directional selection for the proinflammatory alleles of the genes that are involved in the humoral immune response mediated by T-helper cells of the second type (Th2 response). It can be assumed

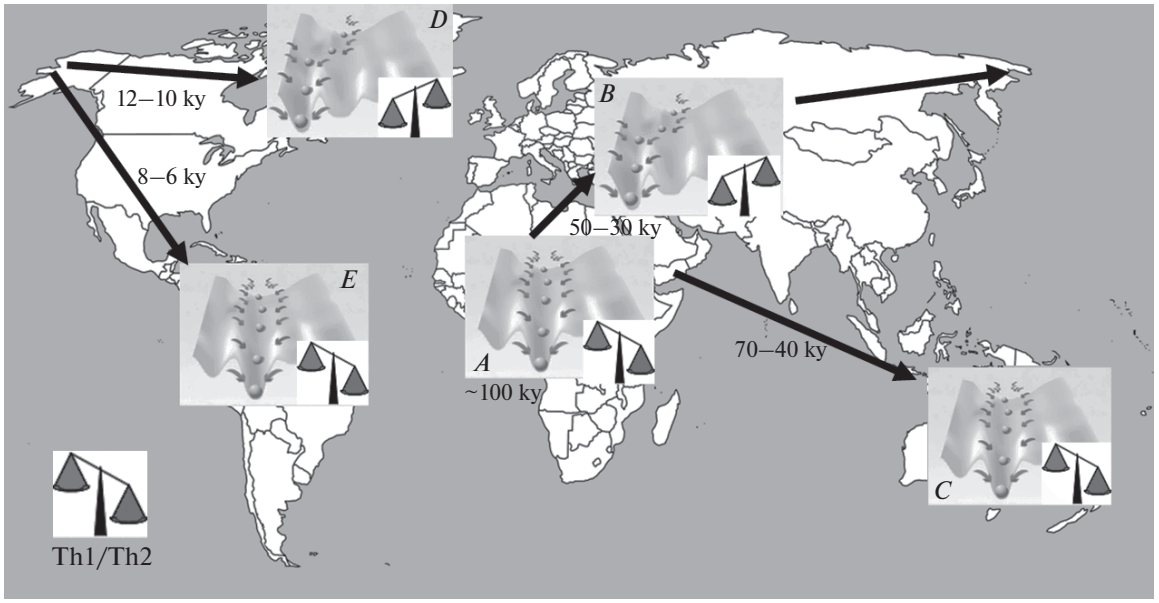

Fig. 3. Hypothesis of canalization and decanalization of the genetic landscape of the immune system in the course of modern human dispersal (by [17]). The genetic landscape is depicted as a terrain of canalized states (valleys, the effect of canalizing environmental factors is shown with arrows). Designations: Th1/Th2—hypothetical balance of the main pathways of the immune response, based on the assumption of the immune system decanalization in the course of human dispersal; ky—thousands of years ago; A—canalized landscape in ancestral and contemporary African populations; B—dispersal to Eurasia and landscape decanalization due to the weakening of the effect of canalizing environmental factors (shift of the Th1/Th2 balance); C— colonization of Australia, preservation of the canalized landscape; D—colonization of America from Asia, preservation of the decanalized landscape in the Inuit; E—colonization of South America. Secondary canalization (recanalization) of the genetic landscape by the initial environmental factors.

that human dispersal to the zones with temperate and arctic climate and the exposure of other infectious agents (viral and intracellular bacterial infections) could cause a shift in the established Th2/Th1 balance and decanalization of the genotype–environment relations established in the ancestral populations.

To test this hypothesis, we studied the distribution of the functional alleles of several genes of the immune system in 15 world populations from different climatic zones of the Earth and estimated the relationship of the genetic diversity to the climatic and geographical factors [20].

In this survey, we identified certain characteristic patterns of the distribution of genetic diversity with respect to the genes of immune diseases in the world populations that distinguish them from the genome-wide diversity structure discussed above. Such patters were termed the "decanalization signals" and are as follows:

(1) Association of a genetic marker with a clinical phenotype or endophenotype (the highly significant association in GWAS was considered as evidence).

(2) Reduction in the frequency of phenotype-associated (ancestral) allele from Africa to Eurasia.

(3) Correlation of genetic diversity with latitude and climate.

(4) Increase in genetic diversity from Africa to Eurasia, in contrast to the genome-wide diversity.

(5) The accumulation of signals of natural selection (positive tests) for "decanalized" markers.

(6) A greater degree of differentiation of the world population by the "decanalized" markers compared to the "neutral" ones.

(7) Possibly, a smaller haplotype diversity in the "decanalized" genome regions compared to the "neutral" ones.

In general, within the framework of the hypothesis of the immune response decanalization during the modern human dispersal, our findings can be summarized as follows (Fig. 3):

(1) In the ancestral population of modern humans in Africa, the genotype–environment relations, which were canalized by the directional selection, ensured an effective immune response to parasitic infections owing to the shift in the immune system balance toward the Th2 pathway. This canalized system has remained in modern Africans and other indigenous populations in the equatorial and tropical climatic zone.

(2) The modern human dispersal from Africa to the zones with temperate and arctic climate was accompanied by a decrease in the intensity of inflammatory reactions, which, according to our hypothesis, indi-

cates a possible shift of the immune response balance toward the Th1 pathway. The environmental factors associated with the adaptation to the equatorial tropical climate and specific infectious load have changed for the environment requiring adaptation to a more severe climate and a greater efficiency of the humoral immune response. The environmental canalization factor has disappeared, and the system was decanalized. In the modern populations, this decanalization can manifest itself as an increased susceptibility to a number of common diseases (asthma, systemic lupus erythematosus, etc.), inherited from the ancestral populations, where this susceptibility was cryptic (hidden).

(3) When the populations that have passed a long way of changes in gene frequencies return to the initial environmental conditions of the tropical climate (e.g., South American Indians), the canalizing environmental factor comes back and is sufficiently strong to return the frequencies of the functionally significant proinflammatory alleles to a state close to the initial one (recanalization).

HUMAN DISEASES AND ADAPTATION TO CLIMATE

The decanalization of the immune response in the course of human dispersal is an example of possible adaptive changes in gene frequencies and genotypes in human populations upon their migration from Africa under directional action of environmental factors (in this case, the change of the infectious load, which shifted the immune response balance), which are associated with the climatic and geographical environmental conditions. These results, obtained within the framework of a candidate approach (selection of genetic markers associated with the phenotype candidate for decanalization), stimulated us to perform a positional genome-wide search for the signals for decanalization and adaptation to climate changes in the course of human dispersal. The questions that we attempted to answers were as follows: What are the genes that could be the targets of natural selection in the human dispersal? What biological processes could be associated with the adaptation to the changing environment in the course of human dispersal? What diseases can result from the release of "cryptic variation" if adaptation is related to decanalization and realization of "ancestral susceptibility" to MDs? Although the data of the analysis performed on 79 world population genotyped for 200000 SNPs are still being processed, some of the results are noteworthy.

We identified approximately 1000 SNPs distributed throughout the genome and belonging to approximately 450 genes that are highly significantly ($p < 1 \times 10^{-8}$ with the FDR correction for the genome-wide analysis) associated with key climatic factors and latitude. The search for the diseases associated with these genes according to the Genetic Association Database [40] revealed approximately 1000 significant gene–disease associations. The distribution of these associations with respect to classes of diseases showed that the groups of diseases that are most enriched in associations include the expected groups of phenotypes, such as cardiovascular disease and related metabolic endophenotypes (20% of all associations) as well as immune and infectious diseases (8%). The unexpectedly highly enriched groups included the associations with the pharmacogenetic traits (21%) as well as neurological and psychiatric disorders (30%). Probably, the first group of phenotypes shows the adaptive response to the gene pool to xenobiotics and changes in the diet structure, and the high proportion of neuropsychiatric diseases may be the consequence of the release of the "cryptic variation," associated with the behavioral phenotypes that facilitated the migration activity and the "search for novelty."

According to the results of our analysis, the key biological processes—the gene ontologies significantly enriched in the group of genes that showed the adaptive evolution signals—are the processes that are clustered into the gene ontology groups such as the regulation of metabolism, response to various endogenous and exogenous substrates, signal transduction, transcriptional regulation, neurology, and behavior.

The genetic diversity for 1000 SNPs that showed highly significant associations with the climatic parameters demonstrates clinal growth from Africa to Eurasia, similar to the trend of the immune-dependent phenotype markers, in contrast to the genome-wide diversity markers (Fig. 4 vs. Fig. 2). The analysis of the allele frequencies of these 1000 SNPs in the world population by the principal component analysis (Fig. 5) showed that their distribution pattern was fundamentally different from the genome-wide pattern shown in Fig. 1 and formed primarily by the selectively neutral factors of population dynamics. In this case, we can see the similarity of the genetic structure of all populations of the tropical climate regions—Africa, Oceania, and South America, distant by the first principal component from the Eurasian populations. The latter, in turn, represent a genetic continuum "extended" from the subtropical populations of Southeast Asia (in the area of low positive values of the first factor) to the arctic populations of North Asia, which are characterized by the lowest negative values of the first principal component (Fig. 5).

The similarity of the genetic characteristics of the populations that are heterogeneous in origin but live under similar climatic conditions in this case can be explained by the passage of the evolutionary pathway from the initial canalized state (retained in Africans) through decanalization (or canalization to a new state) in Eurasia and recanalization (return to the initial evolutionary "canal") in the tropics of South America and Oceania.

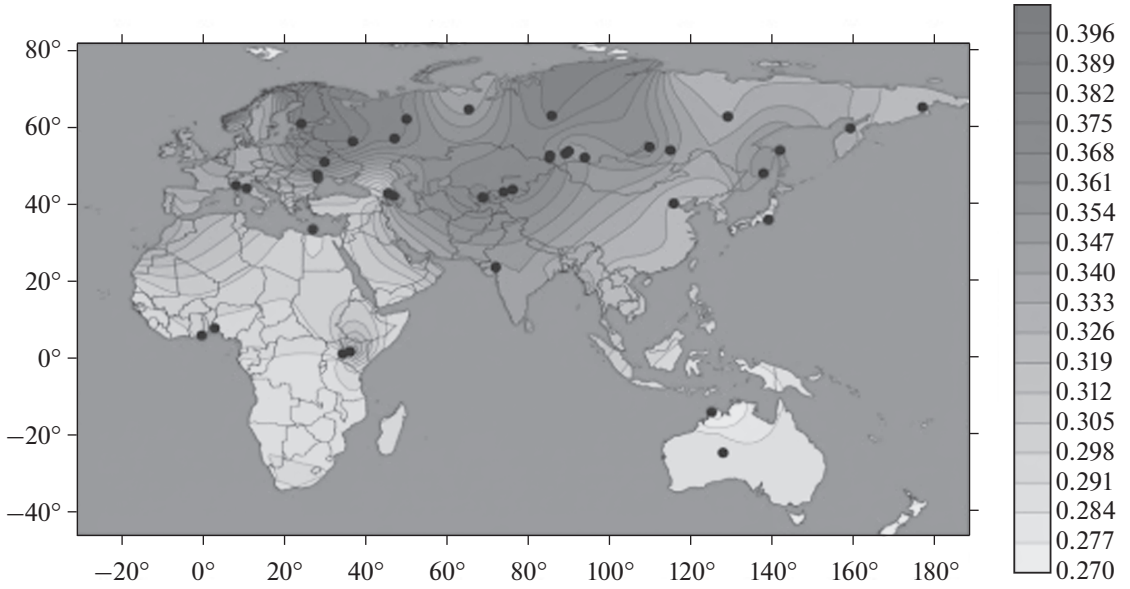

Fig. 4. Genetic diversity (mean heterozygosity) by 1000 SNPs associated with the key climatic factors and demonstrating decanalization signals. The black dots represent the populations for which the mean expected heterozygosity was determined. The scale shows the expected heterozygosity values.

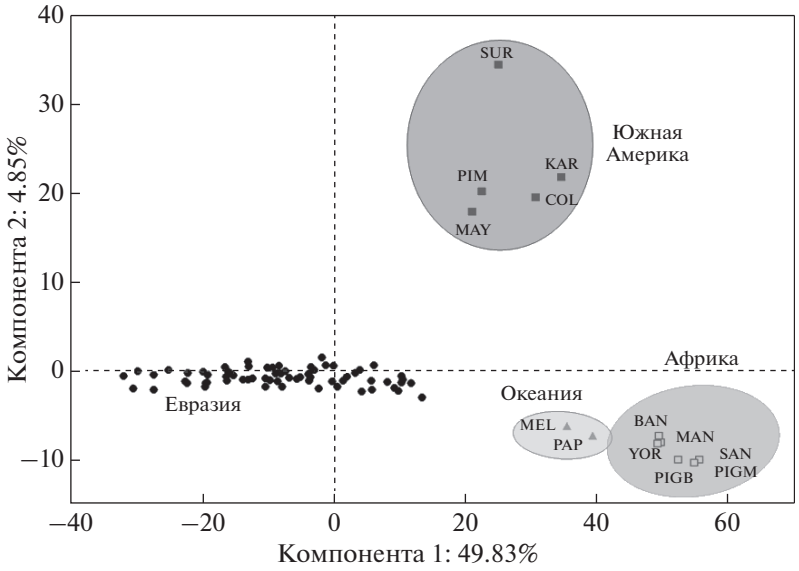

Fig. 5. Space of the principal components of the frequency of 1000 SNPs associated with the key climatic factors and demonstrating decanalization signals in 79 world populations. @Key: Компонента –> Component; Евразия –> Eurasia; Африка –> Africa; Азия –> Asia; Океания –> Oceania; Южная Америка –> South America.

Numerous associations of the genes carrying the "decanalized" markers with multifactorial diseases, apparently, are evidence of changes in the adaptive significance of ancestral alleles (according to the "ancestral susceptibility" concept) and the release of "cryptic variation" in the decanalized Eurasian populations (according to the decanalization concept). The role of natural selection in forming systematic changes in allele frequencies upon human dispersal (decrease in the ancestral allele frequency and increase in heterozygosity) for this group of genes was, probably, significant. Up to 20% of markers showed positive signals of natural selection for individual SNPs. It should be noted that the intensity of selection in the populations from the equator to the poles increased, which was not observed in other genome regions, whose variation was not correlated with the climatic and geographical parameters of the environment. The analysis of selec-

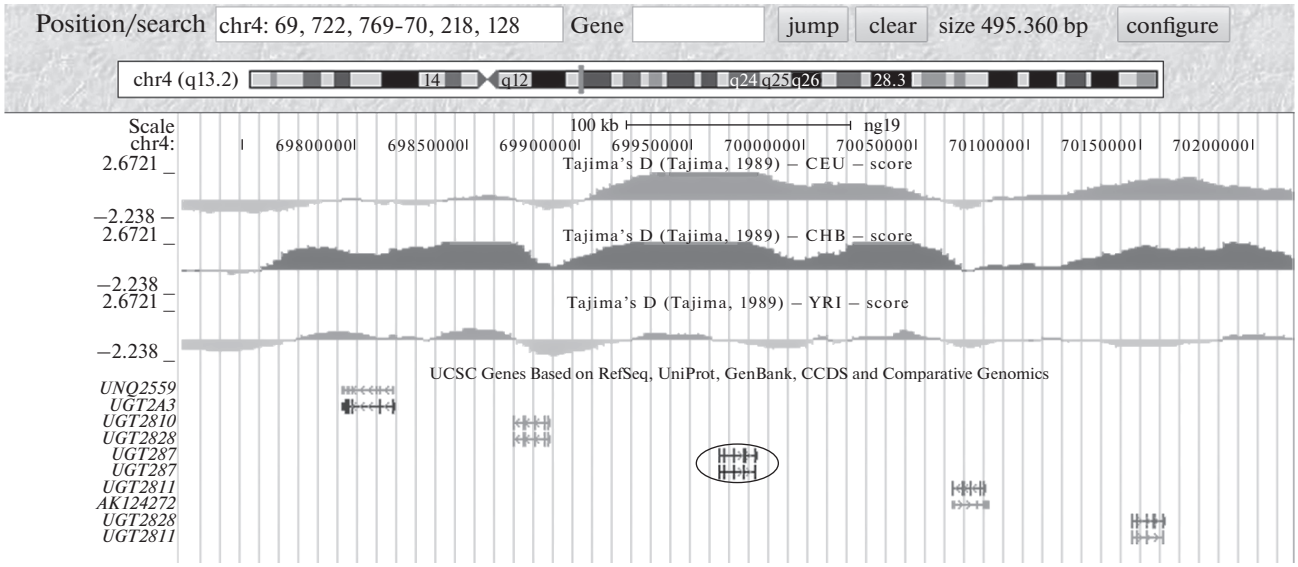

Fig. 6. Balancing selection signal in the chromosome 4 region in the genome browser of the 1000 Genomes project. The ~500000-bp region comprises the *UGT2B7* gene and four other protein-coding genes. The Tajima test values along the chromosome are shown for Europeans (CEU, upper plot), Asians (CHB, middle plot), and Africans (YRI, lower plot). The localization of genes in this chromosome region is shown below. The position of the *UGT2B7* gene is enclosed with an oval.

tion signals in the decanalized genome regions using the tests for the structure of linkage disequilibrium and haplotype variation according to the 1000 Genomes Project identifies valid signals of the balancing and directional selection. In the analysis, we used the Tajima test, the cross-population extended haplotype homozygosity (XP-EHH) test, and the cross-population multilocus composite likelihood ratio (XP-CLR) test.

An example of the effect of selection on the genetic diversity of the loci included in our "top 1000" decanalized marker list is the *UGT2B7* gene, which encodes one of the variants of uridyl-5-diphosphate glucuronosyltransferase, an enzyme involved in xenobiotic detoxification phase II. The function of this protein is the transfer of glycosyl groups to UDP molecules, as a result of which toxic substances are converted to harmless water-soluble glucuronides [41]. Polymorphisms for this gene are associated with diseases such as systemic lupus erythematosis, HIV, obesity, and epilepsy, as well as with the response to drug therapy of a number of diseases [42, 43].

Figure 6 shows a significant signal of the balancing selection for the *UGT2B7* gene on chromosome 4 and in adjacent regions in the European (upper plot) and Asian (middle plot) populations, but not in Africans (lower plot). The selection zone extends to an approximately 150 000-bp area comprising the entire coding region of this gene and its 5'-UTR and 3'-UTR regions. The distribution pattern of the *rs4587017* alleles of the *UGT2B7* gene, showing the most well-expressed signals of selection and decanalization (Fig. 7), is indicative of directional changes in the allele frequencies from the high frequency of the ancestral allele (shown in light gray), which is associated with a slow metabolism, to a high frequency (over 60%) of the derived allele, which is associated with fast metabolism in Northeast Asia. With the human migration back to tropics, the ancestral allele again occurs more frequently (up to its complete fixation in the equatorial populations of Indians in Brazil). That is, here we see the same chain of events mediated by the adaptation to the climate owing to the influence of selection as in the case of the genes associated with immune-related diseases: canalization in ancestral and contemporary African populations, decanalization upon the migration to the north, and recanalization upon the return to the original environmental conditions. In this case, the slow directional selection against the initially most frequent ancestral allele with increasing distance from Africa leads to a gradual increase in the genetic diversity in the majority of nontropical populations, thereby mimicking the balancing selection signal.

CONCLUSIONS

The genetic diversity of modern human populations was formed under the influence of various microevolutionary factors acting in the course of human dispersal over the Earth. The most pronounced were the effects of migrations (isolation by distance) and drift, which formed the picture of a strict geographical structuring of human genetic variation. The effects of natural selection are usually hidden or erased by the stronger effects of genetic drift. However, the modern arsenal of genome-wide methods of analysis of genetic polymorphism, together with

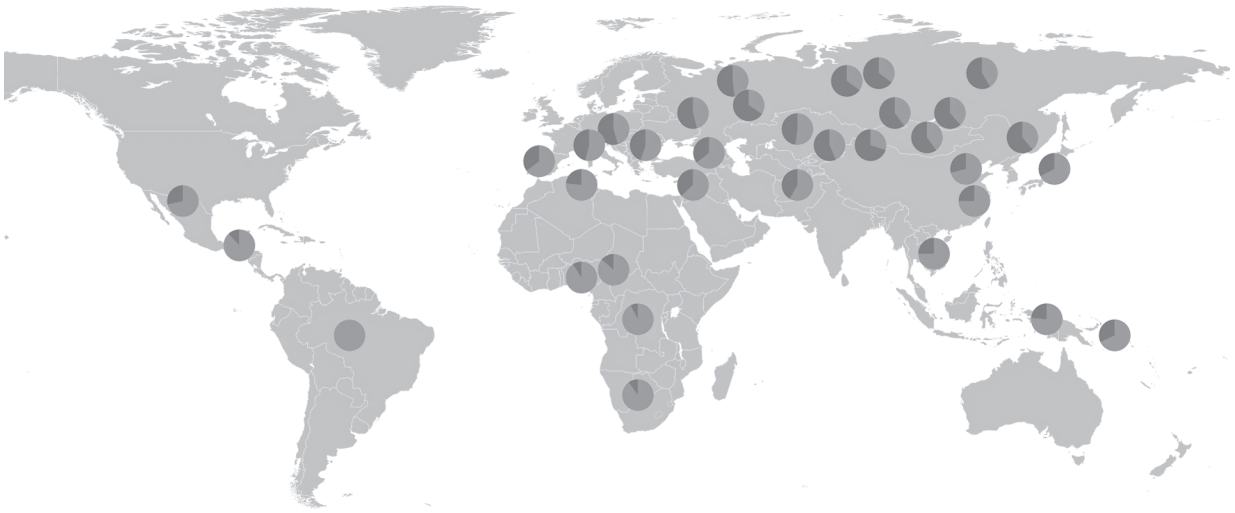

Fig. 7. Frequencies of *rs4587017* alleles of the *UGT2B7* gene in the world populations. The ancestral and derived alleles are shown in light and dark gray, respectively.

the bioinformatics approaches, will make it possible to identify the selection signals and analyze the role of microevolutionary factors in the distribution of common human diseases. This problem is one of the main problems of evolutionary medicine, a currently forming direction in modern biomedicine which studies diseases and human health in the evolutionary context.

Some hypotheses and concepts of evolutionary medicine may be productive for identifying the mechanisms of occurrence, distribution, and pathogenetics of MDs. One of such concepts is the hypothesis of canalization–decanalization of genome–phenome relations in the course of modern human dispersal.

Genome-wide search made it possible to reveal genetic markers and loci of the genome carrying the signals of decanalization and climate adaptation. Part of the human genome variability associated with environmental changes in the course of dispersal from Africa is enriched in the variants associated with neuropsychiatric diseases and behavioral traits, immune and infectious diseases, and response to xenobiotics. The key biological processes that involve these genes and markers are metabolic regulation, signal transduction, response to external stimuli, and nervous system regulation. One of the numerous examples of the genes that carry the selection and decanalization signals is the gene encoding uridyl-5-diphosphate glucuronosyltransferase (UGT2B7), which is involved in the metabolism of xenobiotics and is associated with a number of complex human diseases. Probably, the high frequency of alleles associated with MDs in some human populations (and, perhaps, partly the high frequency of multifactorial diseases themselves) can be explained by the hypothesis of decanalization of genome–phenome relations under the influence of natural selection in the course of human dispersal.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 15-04-002442).

REFERENCES

1. http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi
2. The 1000 Genomes Project Consortium, An integrated map of genetic variation from 1092 human genomes, *Nature,* 2012, vol. 491, pp. 56—65.
3. Manolio, T.A., Collins, F.S., Cox, N.J., et al., Finding the missing heritability of complex diseases, *Nature,* 2009, vol. 461, no. 7265, pp. 747—753.
4. Lee, S.H., Wray, N.R., Goddard, M.E., and Visscher, P.M., Estimating missing heritability for disease from genome-wide association studies, *Am. J. Hum. Genet.,* 2011, vol. 88, no. 3, pp. 294—305.
5. Zuk, O., Hechter, E., Sunyaev, S.R., and Lander, E.S., The mystery of missing heritability: genetic interactions create phantom heritability, *Proc. Natl. Acad. Sci. U.S.A.,* 2012, vol. 109, no. 4, pp. 1193—1198.
6. Knox, S.S., From 'omics' to complex disease: a systems biology approach to gene—environment interactions in cancer, *Cancer Cell Int.,* 2010, vol. 10:11.
7. Chen, L. and Wu, J., Systems biology for complex diseases, *J. Mol. Cell Biol.,* 2012, vol. 4, pp. 125—126.
8. Li, H., Systems biology approaches to epidemiological studies of complex diseases, *WIREs Syst. Biol. Med.,* 2013, vol. 5, pp. 677—686.
9. Puzyrev, V.P. and Kucher, A.N., Evolutionary ontogenetic aspects of pathogenetics of chronic human diseases, *Russ. J. Genet.,* 2011, vol. 47, no. 12, pp. 1395—1405.
10. Stepanov, V.A., Genomes, populations and diseases: ethnic genomics and personalized medicine, *Acta Nat.,* 2010, no. 4, pp. 18—34.

11. Jobling, M., Hollox, E., Hurles, M., et al., *Human Evolutionary Genetics,* New York: Garland Science, 2014, 2nd ed.
12. Meyer, M., Kircher, M., Gansauge, M.-T., et al., A high-coverage genome sequence from an archaic Denisovan individual, *Science,* 2012, vol. 338, pp. 222—226.
13. Prufer, K., Racimo, F., Patterson, N., et al., The complete genome sequence of a Neanderthal from the Altai Mountains, *Nature,* 2014, vol. 505, pp. 43—49.
14. Noonan, J.P., Neanderthal genomics and the evolution of modern humans, *Genome Res.,* 2010, vol. 20, pp. 547—553.
15. Stepanov, V.A., *Ethnogenomics of North Eurasian Population,* Tomsk: Pechatnaya Manufaktura, 2002.
16. Balaresque, P.L., Ballereau, S.J., and Jobling, M.A., Challenges in human genetic diversity: demographic history and adaptation, *Hum. Mol. Genet.,* 2007, vol. 16, rev. issue 2, pp. R134—R139.
17. Wiuf, C., Do DF508 heterozygotes have a selective advantage? *Genet. Res.,* 2001, vol. 78, pp. 41—47.
18. Adeyemo, A. and Rotimi, C., Genetic variants associated with complex human diseases show wide variation across multiple populations, *Public Health Genomics,* 2010, vol. 12, pp. 72—79.
19. Jobling, M.A., Hurles, M.E., and Tyler-Smith, C., *Human Evolutionary Genetics: Origins, Peoples and Diseases,* New York: Garland Publ., 2004.
20. Stepanov, V.A., Candelaria, P., Kho, S., et al., Decanalization of immune response during the dispersion of modern humans: the relationship of genetic diversity in the genes of the immune system with climatic and geographical factors, *Med. Genet.,* 2013, vol. 12, no. 4, pp. 8—18.
21. Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K., A map of recent positive selection in the human genome, *PLoS Biol.,* 2006, vol. 4, pp. 446—458.
22. Knowles, D.G. and McLysaght, A., Recent de novo origin of human protein-coding genes, *Genome Res.,* 2009, vol. 19, pp. 1752—1759.
23. Hancock, A.M., Witonsky, D.B., Alkorta-Aranburu, G., et al., Adaptations to climate-mediated selective pressures in humans, *PLoS Genet.,* 2011, vol. 7, pp. 1—16.
24. Waddington, C.H., Canalization of development and the inheritance of acquired characters, *Nature,* 1942, vol. 150, pp. 563—565.
25. Schmalhausen, I.I., Faktory evolyutsii (teoriya stabiliziruyushchego otbora) (Factors of Evolution: the Theory of Stabilizing Selection), Moscow: Akad. Nauk SSSR, 1946.
26. Schmalhausen, I.I., Factors of Evolution: the Theory of Stabilizing Selection, Philadelphia: Blakiston, 1949.
27. Altukhov, Yu.P., *Geneticheskie protsessy v populyatsiyakh* (Genetic Processes in Populations), Moscow: Akademkniga, 2003, 3rd ed.
28. Altukhov, Yu.P., Korochkin, L.I., and Rychkov, Yu.G., Hereditary biochemical diversity in evolution and development, *Russ. J. Genet.,* 1996, vol. 32, no. 11, pp. 1256—1275.
29. Neel, J.V., Diabetes mellitus: a "trifty" genotype rendered detrimental by "progress"? *Am. J. Hum. Genet.,* 1962, vol. 14, pp. 353—362.
30. Neel, J.V., The "trifty" genotype in 1998, *Nutr. Rev.,* 1998, vol. 57, pp. S2—S9.
31. Di Rienzo, A., and Hudson, R.R., An evolutionary framework for common diseases: the ancestral-susceptibility model, *Trends Genet.,* 2005, vol. 21, no. 11, pp. 596—601.
32. Gibson, G., Decanalization and the origin of complex disease, *Nat. Rev. Genet.,* 2009, vol. 10, no. 2, pp. 134—140.
33. McGrath, J.J., Hannan, A.J., Gibson, G., Decanalization, brain development and risk of schizophrenia, *Transl. Psychiatry,* 2011, vol. 1. e14.
34. Scharloo, J., Canalization: genetic and developmental aspects, *Annu. Rev. Ecol. Syst.,* 1991, vol. 22, pp. 65—93.
35. Siegal, P. and Bergman, J., Waddington's canalization revisited: developmental stability and evolution, *Proc. Natl. Acad. Sci. U.S.A.,* 2002, vol. 99, no. 16, pp. 10528—10532.
36. Bethony, J., Brooker, S., Albonico, M., et al., Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm, *Lancet,* 2006, vol. 367, pp. 1521—1532.
37. WHO Expert Committee, Prevention and control of schistosomiasis and soil-transmitted helminthiasis, in *World Health Organ Tech. Rep. Ser.,* Geneva, 2002.
38. Hughes, R.G., Environmental influences on helminthiasis and nutritional status among Pacific schoolchildren, *Int. J. Environ. Health Res.,* 2004, vol. 14, pp. 163—177.
39. Le Souef, P.N., Goldblatt, J., and Lynch, N.R., Evolutionary adaptation of inflammatory immune responses in human beings, *Lancet,* 2000, vol. 356, pp. 242—244.
40. http://geneticassociationdb.nih.gov/
41. Mackenzie, P., Little, J.M., and Radominska-Pandya, A., Glucosidation of hyodeoxycholic acid by UDP-glucuronosyltransferase 2B7, *Biochem. Pharmacol.,* 2003, vol. 65, pp. 417—421.
42. Fujita, K.I., Ando, Y., Yamamoto, W., et al., Association of UGT2B7 and ABCB1 genotypes with morphine-induced adverse drug reactions in Japanese patients with cancer, *Cancer Chemother. Pharmacol.,* 2009, vol. 65, pp. 251—258.
43. Rouguieg, K., Picard, N., Sauvage, F.L., et al., Contribution of the different UDP-glucuronosyltransferase (UGT) isoforms to buprenorphine and norbuprenorphine metabolism and relationship with the main UGT polymorphisms in a bank of human liver microsomes, *Drug Metab. Dispos.,* 2010, vol. 38, pp. 40—45.

Translated by M. Batrukova