

GENOMICS.
TRANSCRIPTOMICS

UDC 599.323.45

Variability of Fragments of Nuclear *Brca1* Gene, Exon 11, and Mitochondrial *Cox1* Gene in House Mice *Mus musculus*

A. S. Bogdanov^{a,*}, A. N. Maltsev^b, E. V. Kotenkova^b, V. G. Malikov^c, A. A. Lisovsky^d,
V. V. Stakheev^e, J. Darvish^{f,†}, and R. Castiglia^g

^aKoltzov Institute of Developmental Biology, Russian Academy of Sciences, Moscow, 119334 Russia

^bSevertsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow, 119071 Russia

^cZoological Institute, Russian Academy of Sciences, St. Petersburg, 199034 Russia

^dZoological Museum, Moscow State University, Moscow, 125009 Russia

^eSouthern Scientific Center, Russian Academy of Sciences, Rostov-on-Don, 344006 Russia

^fFerdowsi University of Mashhad, Mashhad, 9177948974 Iran

^gUniversità degli Studi di Roma “La Sapienza”, Rome, 00161 Italy

*e-mail: bogdalst@yahoo.com

Received June 6, 2019; revised September 27, 2019; accepted October 7, 2019

Abstract—To clarify genetic differences between subspecies of the house mouse *Mus musculus*, their distribution, and hybridization, we first conducted a comparative analysis of variability of nucleotide sequences of fragments of the nuclear gene *Brca1*, exon 11 (2331 bp), and mitochondrial gene *Cox1* (1260 bp) in 40 house mice from West and East Europe, Transcaucasia, Siberia, and Central and South Asia. *Brca1* genotypes were divided into five main groups, which differed in a number of fixed substitutions. Genotypes of each group are characteristic for the certain geographical region and the following subspecies: *M. m. musculus*, *M. m. domesticus*, *M. m. castaneus*, and *M. m. wagneri* together with *M. m. gansuensis*; a fifth group corresponds to an unidentified subspecies or a distinct genetic form of *M. musculus* from India (Sikkim State). Besides the homozygous specimens, we revealed mice, which were heterozygous for all diagnostic loci simultaneously; these specimens were determined as hybrid. Hybrid mice were mainly found in the zones of contact of subspecies, but in some cases, quite far from one of the parent subspecies (possibly, due to transportation). In two hybrid mice (from Bakhtiari Province of Iran and Transbaikalia of Russia), unique *Brca1* haplotypes were detected. It cannot be ruled out that, at least partly, they may be characteristic of the *M. m. bactrianus* and *M. m. gansuensis* subspecies, respectively. Thus, the results of the study showed that the nuclear *Brca1* gene is a promising molecular genetic marker for the analysis of variability, differentiation, and hybridization of house mice as well for subspecific identification of *M. musculus* specimens. Despite more rapid evolution of the *Cox1* gene, it is not well suited for discrimination of *M. m. musculus*, *M. m. wagneri*, *M. m. gansuensis* specimens and Transcaucasian representatives of *M. m. domesticus* due to introgression and long-term maintenance of foreign mitochondrial DNA in populations. However, *Cox1* gene analysis (along with the diagnostics of animals by nuclear DNA) may be useful for estimation of population differences in *M. m. castaneus* and *M. m. domesticus* subspecies.

Keywords: *Mus musculus*, exon 11 of *Brca1* gene, *Cox1*, intraspecific variability and differentiation

DOI: 10.1134/S0026893320020028

INTRODUCTION

House mice (genus *Mus*) are a heterogeneous group that includes numerous forms in varying degrees differing from one another in morphological, ecological, ethological features and genetic characteristics. Due to the high diversity, asynchronous variability of morphological and genetic traits, and due to hybridization of synanthropic forms, which considerably complicates distinguishing discrete population groups and determination of their specific characteris-

tics, the taxonomic system of the genus has been repeatedly revised and currently is not generally accepted. Only the species level of the differences between the “wild” house mice in relation to the synanthropic house mice can be recognized as indisputable. The “wild” forms (the Macedonian mouse *M. macedonicus*, the steppe mouse *M. spicilegus*, and the Western Mediterranean mouse *M. spretus*) are also currently identified by the majority of researchers [1–3] as allospecies; i.e., their status in fact is equal to the species.

Numerous parapatric synanthropic forms are combined in the house mouse species, *M. musculus* (or

[†] Deceased.

superspecies, from a different point of view). Their taxonomy remains controversial despite the fact that *M. musculus* is one of the main laboratory objects of genetic and molecular biological research, and its complete genome was sequenced second after the human genome [4]. Typically, the synanthropic forms of the house mouse are considered as subspecies [2, 3, 5–8], although it has been proposed to rise the taxonomic rank of some of them (*M. m. domesticus*, *M. m. castaneus*, and *M. m. musculus*) to species [9–13]. Until recently, contradictory information was also published about the Central Asian subspecies *M. m. wagneri* and Transbaikalian *M. m. gansuensis*, which were considered morphologically and genetically closest to *M. m. musculus*. For example, according to Spiridonova et al. [14], all *M. musculus* subspecies can be identified by the complex of RAPD-PCR markers, and the level of differences between some of them (*M. m. wagneri* and *M. m. musculus*) is comparable to interspecies differences. In addition, successful identification of individuals of *M. m. wagneri*, *M. m. gansuensis*, and *M. m. musculus* by karyotypic characters was reported [15]. However, the analysis of variability of the second intron of the nuclear hemoglobin b1 chain gene (*Hbb-b1*) revealed only two genotypes in the house mice of Central Asia and Transbaikalia, one of which was identified as characteristic of *M. m. musculus*, and the other was found to be common for *M. m. wagneri* and *M. m. gansuensis* [16]. Likewise, in the study of polymorphism of eight linked genes of the eighth autosome, two phylogroups of house mice (“musculus” and “wagneri”) were identified in Asia [17]. The results of recent studies of mtDNA fragments—the control region (*D-loop*) and the cytochrome *b* gene (*Cytb*)—showed a complete, approximately the same, though noticeably inferior to the allopecies of “wild” mice, genetic differentiation of the subspecies *M. m. domesticus*, *M. m. musculus*, *M. m. castaneus*, *M. m. gentilulus*, *M. m. bactrianus*, a new subspecies *M. m. isatissus*, and an unclassified form from Nepal [7, 8, 16, 17]. However, despite the rather high variability of the mitochondrial genes, the subspecies *M. m. musculus*, *M. m. wagneri*, and *M. m. gansuensis* did not form individual clades on dendrograms [7, 16–18]. Distinguishing subclusters, which often included mice originating from very distant regions and belonging to different subspecies, had an apparently random nature. It was assumed that the mtDNA in the subspecies *M. m. wagneri* and *M. m. gansuensis* was substituted with the mtDNA of *M. m. musculus* as a result of “absorbing” hybridization [16]. Due to the limited genetic evidence in favor of isolation of *M. m. wagneri* and *M. m. gansuensis* with respect to both each other and *M. m. musculus*, in the checklist on mammals of the world fauna [3] the first two subspecies are recognized as synonymous to the last subspecies.

The study of the diversity, phylogenetic relationships, and hybridization of house mice is still based on the use of primarily mitochondrial genes, which

evolve much faster than nuclear genes [19–21]. Apparently, for this reason, the nuclear genes were studied (except for several mentioned papers) when comparing species of house mice and the most distinct *M. musculus* subspecies in a limited number of animals [13, 22–24]. However, the analysis of mtDNA, which has a different mode of inheritance (is transmitted to descendants through the egg cytoplasm, i.e., by the maternal line) and recombines neither with other mtDNA molecules nor with nDNA, can reveal a mixed composition of the population but does not make it possible to identify a hybrid origin of individuals without using nuclear markers. Thus, the discrepancy between the results of identification by karyotypic, nuclear markers, on the one hand, and mitochondrial markers, on the other hand, is not surprising and may even be expected, especially in hybrid zones. In this regard, it seems necessary to analyze the most rapidly evolving nuclear autosomal genes. Therefore, the selection of a suitable nuclear marker exhibiting specificity in all or at least in the majority of intraspecific forms of *M. musculus* (which is required for studying hybridization) and population variability to analyze genetic polymorphism is of paramount importance.

The main purpose of this study was to investigate the genetic diversity of house mice from several populations of Europe, Transcaucasia, Central Asia, and Siberia by the nucleotide sequence of a fragment of exon 11 of the autosomal nuclear gene *Brcal* (Breast Cancer 1) as well as to assess the suitability of this marker for the analysis of phylogenetic relationships and hybridization of intraspecific forms of *M. musculus*. In addition to the *Brcal* gene, in some individuals we sequenced a fragment of the mitochondrial gene encoding the first subunit of cytochrome oxidase (*Cox1*) to determine the level of polymorphism of this marker in the group of house mice, to refine the genetic differences identified as a result of analysis of the *Brcal* gene, and to additionally verify the hybrid origin of some specimens.

MATERIALS AND METHODS

In total, we studied 40 *M. musculus* individuals belonging to at least five subspecies: *M. m. musculus*, *M. m. wagneri*, *M. m. gansuensis*, *M. m. domesticus*, and *M. m. castaneus*. Two Macedonian mice (*M. macedonicus*) and one steppe mouse (*M. spicilegus*) were used as outgroups. The localities where mice were caught, the composition of samples in which *Brcal* and *Cox1* genes were studied, and the accession numbers of the nucleotide sequences deposited in the GenBank database are shown in Table 1. In addition to our own material, the previously published [25–28] sequences of the *Brcal* gene of laboratory mouse strains C57BL/6 and C57BL/6j (GenBank accession numbers: XM_017314233, BC068303, NM_009764, U35641, and U31625), which originated from the

Table 1. Material of our own collections used in the study

Species	Collection number	Capture locality	GenBank accession number	
			<i>Brca1</i>	<i>Cox1</i>
<i>M. musculus</i>	SBP1	1. Italy, Granica	MN228628	MN228588
	SBP14	2. Italy, Collebaccaro	MN228629	MN228589
	26258	3. Ukraine, vicinities of Kyiv	MN228603	MN228581
	26637	4. Belarus, Minsk oblast, Volozhinskii raion, vicinities of villages Sakovschina and Kaldyki	MN228622	–
	26643		MN228616	–
	26920	5. Russia, Moscow oblast, Noginsk raion, vicinities of Ivanovskoe village	MN228604	–
	26921		MN228617	–
	24133	6. Russia, Kursk oblast, vicinities of Kursk, Zapovednyi village	MN228598	MN228578
	26071	7. Russia, Krasnodar krai, Temryuk raion, vicinities of Primorskii village	MN228615	–
	26072		MN228602	–
	3	8. Russia, Astrakhan oblast, 45 km southwards of Astrakhan, Volga River delta	MN228605	–
	5		MN228609	–
	100	9. Armenia, Yerevan	MN228630	–
	126		MN228631	MN228590
	240		MN228632	MN228591
	24980	10. Azerbaijan, Lenkoran raion, vicinities of Dashtatyuk village (former Alekseevka)	MN228634	–
	24981		MN228614	–
	406	11. Kazakhstan, West Kazakhstan region, Esenankaty River floodplain	MN228606	–
	413	12. Kazakhstan, West Kazakhstan region, Akzhaiksky raion, Taipakskii rural district, Shauesh wintering site	MN228607	–
	24634	13. Russia, Tyumen oblast, Yamalo-Nenets autonomous district, Labytnangi	MN228600	MN228579
	24635		MN228601	MN228580
	16	14. Russia, Tyumen oblast, Ishim	MN228633	–
19	MN228623		–	
24223	15. Russia, Novosibirsk oblast, vicinities of Novosibirsk near Akademgorodok, 1 km northwards of Kirovo village	MN228610	–	
24227	16. Russia, Altai krai, Soltonskii raion, Solton village	MN228611	–	
24228		MN228599	–	
418	17. Russia, Transbaikalia, Chita	MN228624	–	
424	18. Russia, Transbaikalia, Onon raion, Bolshevik village, Daurskii Reserve	MN228625	MN228586	
425		MN228626	MN228587	
437	19. Russia, Transbaikalia, Onon raion, vicinities of Nizhnii Tsasuchei village, Tsasuchei forest	MN228627	–	
447	20. Russia, Transbaikalia, Onon raion, Ust-Imalka village	MN228608	MN228582	
24245	21. Kazakhstan, East Kazakhstan region, Tarbagatai raion, western coast of Lake Zaisan	MN228619	–	
24279		MN228612	–	
24295	22. Kazakhstan, Almaty region, Balkhash raion, 8 km from Bakanas village towards Akkol village	MN228613	–	

Table 1. (Contd.)

Species	Collection number	Capture locality	GenBank accession number	
			<i>Brca 1</i>	<i>Cox 1</i>
<i>M. musculus</i>	24330	23. Kazakhstan, Kyzylorda region, Chiiliiskii raion, left bank of the Syrdar'ya River, 1.5 km downstream of Tartogai village	MN228620	MN228584
	24383	24. Turkmenistan, Balkanskii province, Syunt-Khasardag Reserve, Kopetdag Ridge	MN228621	MN228585
	25559	25. Tajikistan, Rogun raion, vicinities of Obi Garm village	MN228618	MN228583
	1214	26. India, Himachal Pradesh State, Kullu Valley, Naggar village	MN228637	MN228594
	1140	27. India, Sikkim State, Maenam Hill	MN228636	MN228593
	26925	28. Iran, Bakhtiari province, Kohne Pist	MN228635	MN228592
<i>M. macedonicus</i>	26922	29. Iran, Qazvin province, Buin Zahra environs,	MN228639	MN228596
	26923	Feyzabad village	MN228640	MN228597
<i>M. spicilegus</i>	K-14	30. Moldova, vicinities of Chisinau, Cricova	MN228638	MN228595

Table 2. Primers used for PCR and sequencing exon 11 of the nuclear gene *Brca 1* and the mitochondrial gene *Cox 1*

Gene	Fragment	Sequences of the forward (upper line) and reverse (bottom line) primers, 5'–3'	Source
<i>Brca 1</i> , exon 11	I	BRCA1-1F (GGCATAAGAAACATTGAACATCATCA)	[32]
		BRCA1-M1R (TTTCTTCACTGCTACCACAACATAT)	
	II	BRCA1-M2F (ACGTCCACAGTTCAAAAGCACCTA)	Selected by us
		BRCA1-M2R (TGAAAATAAAGCAAATGACTGAC)	
	III	BRCA1-M3F (GTGTATGACTCAGTTTATCGC)	[32]
		BRCA1-3R (TTAGGCCCTCTGTTTCTACCTAG)	
<i>Cox 1</i>	I	L5310 (CCTACTCAGCCATTTTACCTATG)	[33]
		R6036 (ACTTCTGGGTGTCCAAAGAATCA)	
	II	L5850-Mus (TAACACAATATCAAACCTCCA)	Selected by us
		R6560-Mus (GAAGAATGTTATGTTTACTCCTA)	

West European house mouse *M. m. domesticus* and whose genomes contain approximately 80% of the genes of this subspecies [29, 30], were included in the analysis.

Total DNA was isolated by the phenol–chloroform method after the treatment of ground tissues with proteinase K [31] or by using a GeneJET™ reagent kit (Fermentas, United States). The nucleotide sequence of the fragment of exon 11 of the *Brca 1* gene was determined by sequencing three overlapping regions separately. When analyzing the *Cox 1* gene, two overlapping regions were sequenced. Preliminary cloning of the fragments of both genes was not performed. The primers used for polymerase chain reaction (PCR) and sequencing are listed in Table 2. PCR was conducted in a mixture containing 25–35 ng of DNA, 2 µL of

10× Taq buffer, 1.6 µL of 2.5 mM dNTP, 4 pM of each primer, 1 unit of *Taq* polymerase, and deionized water to a final volume of 20 µL. Amplification was carried out in a Tertsik thermal cycler (Russia) and included preheating at 94°C (3 min) and then 35 cycles in the following mode: 30 s at 94°C; 1 min at 55–57°C (for amplification of fragments II and III of the *Brca 1* gene and fragment II of the *Cox 1* gene), 60°C (for amplification of fragment I of the *Brca 1* gene), or 63°C (for amplification of fragment I of the *Cox 1* gene); 1 min at 72°C; and final elongation of the PCR products at 72°C (6 min). Automatic sequencing was performed using the ABI PRISM®BigDye™ Terminator v. 3.1 kit (ABI, United States) in the AB 3500 genetic analyzer (ABI, United States) at the Koltzov Institute of Developmental Biology, Russian Academy of Sciences.

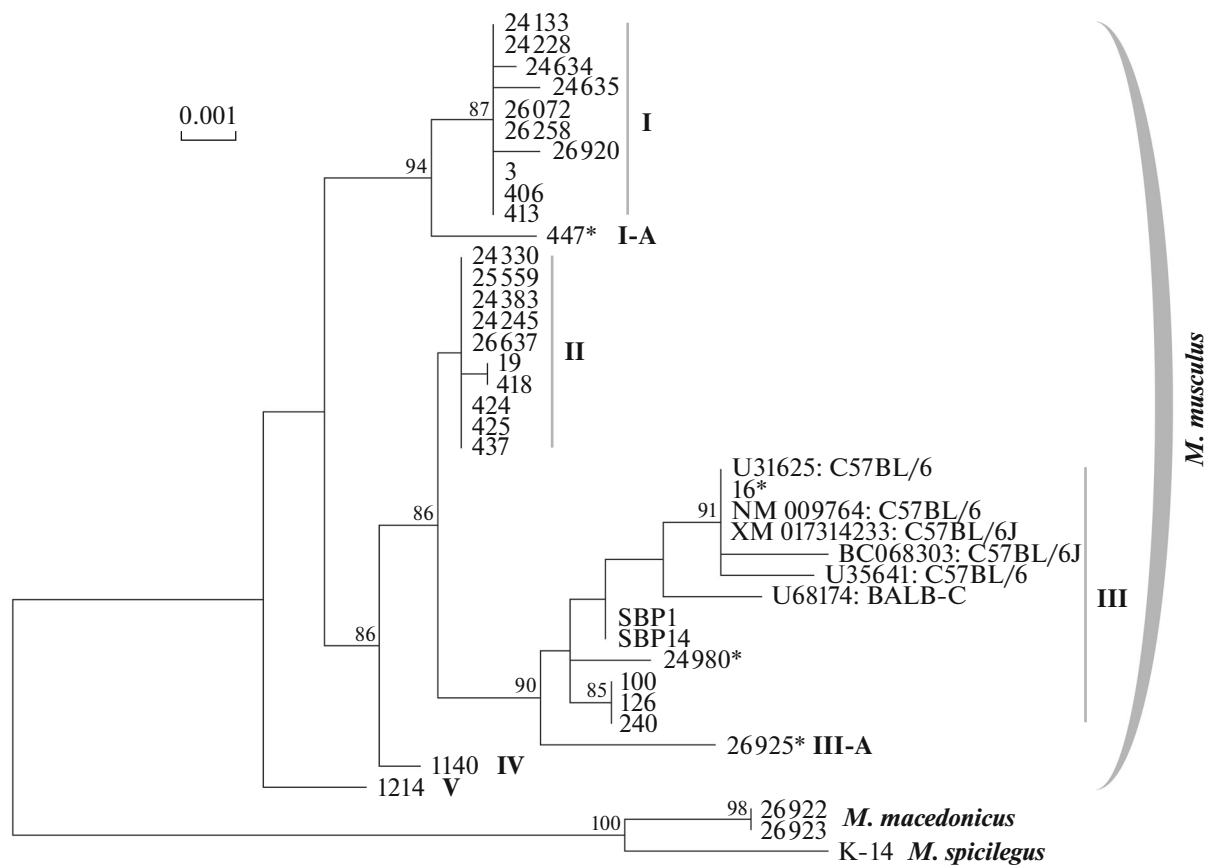


Fig. 1. The ML-dendrogram constructed as a result of comparison of the exon 11 fragment of the *Brca1* gene (2331 bp) of individuals of *M. musculus*, *M. spicilegus*, and *M. macedonicus* from natural populations as well as laboratory mice of C57BL/6 and C57BL/6j lines. For phylogenetic reconstruction we used all specimens homozygous and heterozygous for single nondiagnostic substitutions. In the second case, of two nucleotides of a heterozygous locus we selected the nucleotide that most frequently occurred in the total sample of mice. The individuals that were heterozygous for a number of nucleotide substitutions (hybrid) were excluded from the analysis, except for specimens nos. 447 and 26925 (in constructing the dendrogram, their hypothetically determined unique haplotypes were used) and nos. 16 and 24980 (group III haplotype was used). The hybrid mice are indicated with asterisks. The bootstrap index values exceeding 70% are indicated at the branching nodes of the dendrogram, and the collection numbers of animals are indicated to the right of the branches.

After the addition and alignment of the nucleotide sequences, 2331-bp and 2328-bp (with deletion) fragments of the *Brca1* gene were obtained and analyzed. The studied region begins at the first nucleotide of codon 254 of the translated part of the gene. This is a protein-coding fragment throughout its length corresponding to a sequence of 776–777 amino acid residues in different individuals. The studied region of the *Cox1* gene consisted of 1260 bp starting with its first codon.

Data were statistically processed in the Mega 6.06 software [34]. A suitable model of nucleotide substitutions was selected according to the Bayesian Information Criterion (BIC). When analyzing the *Brca1* gene fragment, we used the TN93 model (Tamura–Nei model) to construct a dendrogram by the Maximum Likelihood (ML) method and calculate genetic distances (D), and the “partial deletion” option was applied to account for the nucleotides that were lost as

a result of deletions as separate characters, similar to the substitutions. When analyzing the *Cox1* gene fragment, we used the T92 + G model (Tamura 3-parameter model with gamma distributed). Bootstrap index values were determined for 1000 replications.

RESULTS

The variability level of the studied fragment of the *Brca1* gene, in general, was high. Significant differences were found not only between the species of house mice, especially between the “wild” (*M. macedonicus* and *M. spicilegus*) mice and synanthropic *M. musculus*, but also between different populations of the latter species and, in some cases, even between individuals of the same population and line, which was quite unexpected for a nuclear gene. Thus, differentiation of the *Brca1* gene in *M. musculus* has a complex multilayered nature.

Table 3. Average genetic distances (D) calculated by comparing the nucleotide sequences of the fragment (2331 bp) of the nuclear gene *Brca1* of house mice of five population groups as well as steppe and Macedonian mice

Species	Group	Genetic distance values								
		<i>M. musculus</i>								<i>M. spicilegus</i>
		I (0.000)	I-A	II (0.000)	III (0.002)	III-A	IV	V	Species as a whole (0.005)	
<i>M. musculus</i>	I	—								
	I-A	0.003	—							
	II	0.005	0.005	—						
	III	0.008	0.009	0.004	—					
	III-A	0.008	0.009	0.005	0.005	—				
	IV	0.004	0.005	0.002	0.005	0.006	—			
	V	0.005	0.006	0.005	0.007	0.007	0.004	—		
<i>M. spicilegus</i>		0.022	0.023	0.023	0.025	0.025	0.021	0.021	0.023	—
<i>M. macedonicus</i>		0.021	0.021	0.021	0.023	0.023	0.020	0.019	0.022	0.005

Average intragroup or intraspecific distances (for samples represented by more than one individual) are shown in parentheses. Distances were calculated using the same samples, genotypes, and haplotypes that were used for constructing the ML-dendrogram (Fig. 1).

At the first, “deeper” level, a clear separation of *M. musculus* genotypes into five groups (I–V) for a number of fixed substitutions is observed. For example, groups I and II differ from each other in six transitions and two transversions. For another two transitions, the “sorting” of genotypes is distinct though incomplete. Group III, in comparison with group II, is characterized by three fixed transitions, and an incomplete “sorting” of genotypes for two transversions is observed. A common feature of the genotypes belonging to groups I and II is the deletion of three nucleotides in a row in the last quarter of the investigated fragment of the *Brca1* gene, which leads to the elimination of one amino acid residue (glutamic acid) from the polypeptide chain and substitution for the upstream residue (phenylalanine with leucine).

In addition to the individuals homozygous for the fixed substitutions in different groups, we also found specimens heterozygous for all marker loci simultaneously. This result indicates the existence of several genetically discrete groups within the species *M. musculus*, and the individuals heterozygous for all marker loci can be identified as hybrid. In the majority of them (nos. 5, 16, 24223, 24227, 24279, 24295, 24980, 24981, 26071, 26643, and 26921), combinations of haplotypes typical for groups I, II, and III were found; however, in several hybrid individuals, unique haplotypes were revealed. Special attention should be given to two mice. Mouse no. 26925 from the Bakhtiari province (Iran) was closest to group III animals; however, in the last quarter of the studied fragment of the *Brca1* gene, it carried five substitutions in heterozygous state, which are unique for the entire *M. musculus* sample. In individual no. 447 from Ust-Imalka village

(Transbaikalia), as well as in the animals from neighboring localities, all marker substitutions characteristic of group II mice can be found. However, in this mouse, seven of the eight substitutions characteristic of group I were detected in the heterozygous state, as well as four unique substitutions and a three-nucleotide insertion characteristic of mice of groups III–V. The unique haplotypes, which were hypothetically determined after the removal of the haplotype typical for groups II or III from the combined “hybrid” sequence of several heterozygous individuals, were then used for subsequent analysis.

The dendrogram presented in Fig. 1 shows the differences and phylogenetic relationships between the species of house mice and population groups of *M. musculus* for the *Brca1* gene including the unique haplotypes found in several hybrids. The latter formed separate branches adjacent to groups I (I-A, ind. no. 447) and III (III-A, ind. no. 26925). It can be assumed with a high probability that the addition of similar sequences to the analysis will result in the formation of additional clusters in the dendrogram. The dendrogram, similarly to the genetic distances ($D = 0.002–0.009$; on average, 0.005) (Table 3) demonstrates the high level of variability of the *Brca1* gene within the species *M. musculus* and the differentiation of its population groups. It is noteworthy that the average genetic distance between them is comparable with the distance between the Macedonian and steppe mice (0.005), and the maximum distance almost twice exceeds the level of differences between *M. macedonicus* and *M. spicilegus*.

The second level of variation and differentiation of *M. musculus* is represented by population differences,

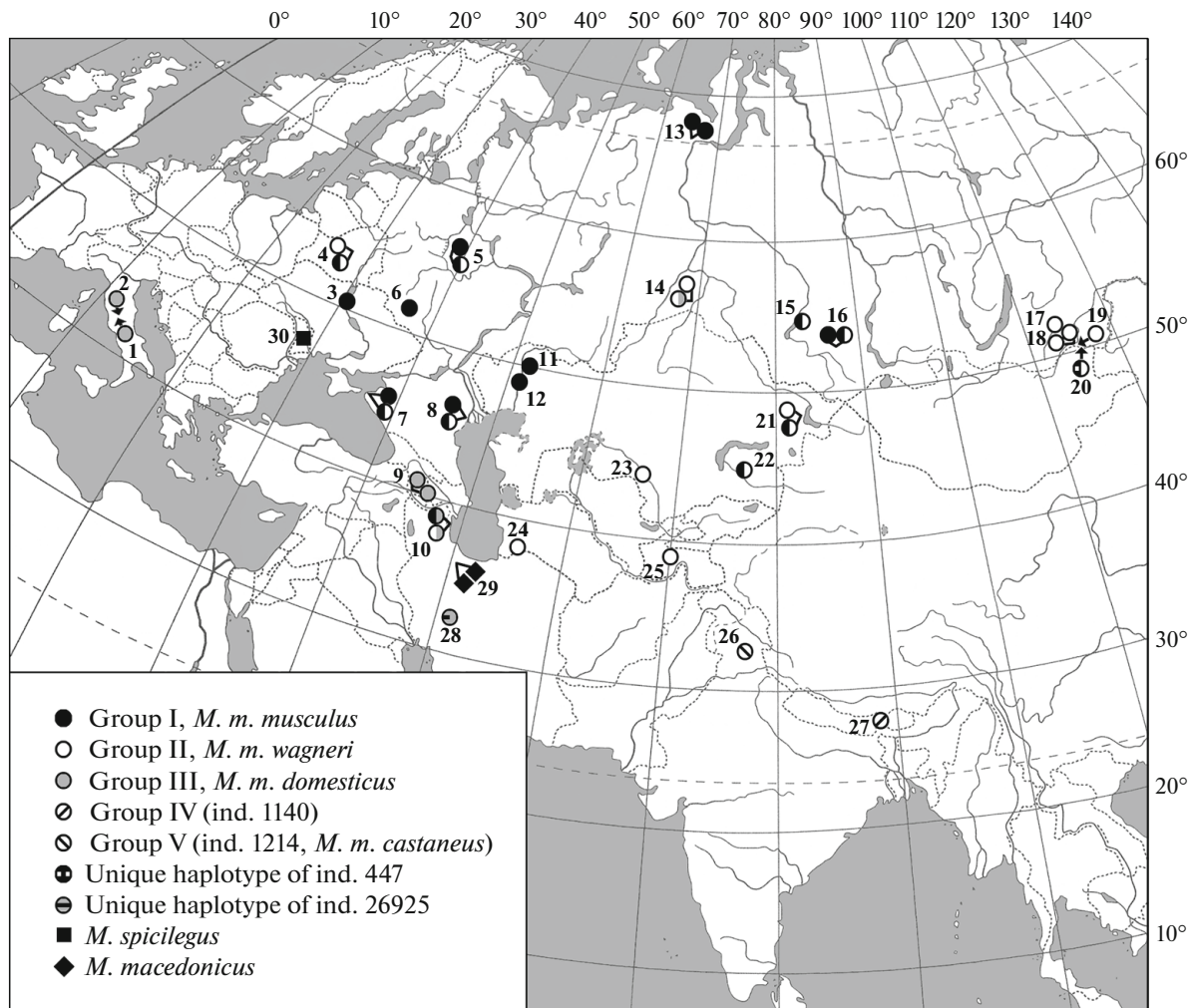


Fig. 2. Distribution of house mice with different *Brca1* genotypes in the studied natural populations. Each symbol corresponds to a single individual. Combined symbols indicate heterozygous (hybrid) mice with different haplotypes. For catchment locality numbers, see Table 1.

especially pronounced in cluster III: the average genetic distance in this group is 0.002. Different genotypes of the *Brca1* gene were detected even in mice of the same line (C57BL/6 or C57BL/6j), as mentioned above. Clusters I and II are more compact, although in one case there is also intrapopulation variability (nos. 24634 and 24635 from Labytningi, group I).

Figure 2 shows the distribution of the identified *Brca1* genotypes in the natural populations of house mice. In general, a clear relation to certain regions can be traced in revealed groups of *M. musculus*. For example, group I genotypes were observed mostly in individuals from populations of Eastern Europe and Northern Trans-Urals; group II genotypes, in mice from Central Asia and Siberia; group III genotypes, in animals from Western Europe and Transcaucasia; and group IV and V genotypes, in mice from India. However, in some cases, animals with a certain genotype or haplotype were caught too far away from its main distribution area. For example, the group II haplotype

was found in mice (mostly hybrid) in some localities of Belarus and Moscow oblast; group III haplotype, in one mouse from Ishim (Tyumen oblast); and group I haplotype, in mice caught in the vicinities of Lakes Balkhash and Zaisan (Eastern Kazakhstan). That is, a significant interpenetration of the population groups of *M. musculus*, identified in the analysis of the *Brca1* gene, is observed.

This might be the cause of the insufficiently successful determination of the subspecies affiliation of house mice on the basis of the mitochondrial genes [7, 16–18]. To test this hypothesis, we analyzed the variability of the fragment of the mitochondrial gene *Cox1* in *M. musculus* individuals identified as homozygous (“pure”) members of groups I–V on the basis of the nuclear gene *Brca1* and originating from the regions remote from the hybridization zones. In the analysis, we additionally used only the hybrid specimens nos. 447 and 26925 with the unique nDNA haplotypes. However, the dendrogram constructed as a result of

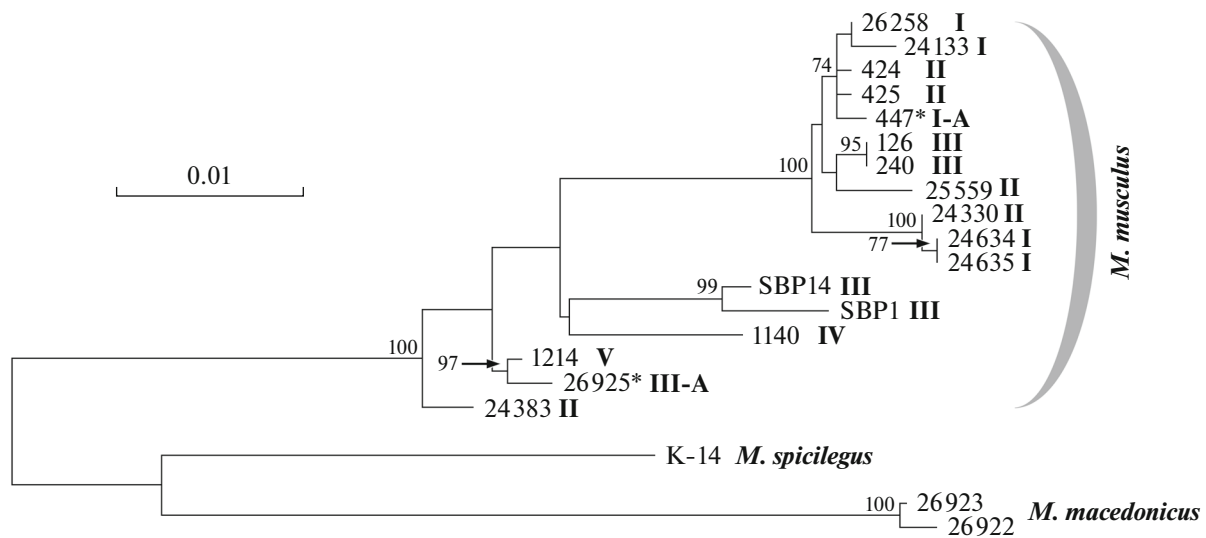


Fig. 3. ML-dendrogram constructed as a result of comparison of the nucleotide sequences of the mitochondrial gene *Cox1* fragment (1260 bp) of individuals of *M. musculus*, *M. spicilegus*, and *M. macedonicus* from the natural populations. The heterozygous (hybrid) mice identified in the study of the *Brca1* gene variability were excluded from the analysis except for individuals nos. 447 and 26925 (indicated with the asterisk). The bootstrap index values exceeding 70% are indicated at the branching nodes of the dendrogram, and the collection numbers of animals and the group to which they belonged according to the *Brca1* gene (Roman numerals) are indicated to the right of the branches.

comparison of the nucleotide sequences of the *Cox1* gene fragment of house mice (Fig. 3) approximately corresponded to the dendrogram obtained as a result of the study of the *Brca1* gene only in terms of interspecies differentiation. Within the species *M. musculus*, the topology of trees, the composition of clusters, and the degree of differences between them were radically different. First of all, in the majority of individuals that, according to *Brca1* gene genotypes, belonged to different intraspecific groups (I, II, and partly III), sufficiently close mitotypes of the *Cox1* gene, uniting in the tree into a single cluster without distinct ordering, were identified. In addition, significant differences in the nature of the phylogenetic relationships of house mice *M. musculus*, especially those at the base of dendrograms, are obvious. Judging by the *Brca1* gene, the basal position was occupied by individuals nos. 1140 and 1214 (belonging to groups IV and V, respectively). However, judging by the *Cox1* gene, the mitotypes of mice nos. 26925 (group III-A) and 24383 (group II) are the most ancient. Finally, judging by the *Cox1* and *Brca1* genes, the level of differences between some individuals was just the opposite. For example, individuals nos. 24634 and 24635 from Labytangi, as mentioned above, differ in their *Brca1* gene genotypes but have the same *Cox1* gene mitotype. The opposite situation is observed in SBP1 and SBP14 mice from Italy, belonging to different chromosomal forms ($2n = 40$ and $2n = 22$, respectively).

DISCUSSION

Sequencing fragments of extended exon 11 of the nuclear gene *Brca1* has been actively performed in

recent years to analyze the phylogenetic relationships of species and genera in different groups of mammals [32, 35, 36]. According to our data, in house mice, the *Brca1* gene shows not only a high interspecific but also intraspecific and even intrasubspecific polymorphism. The pattern of variability and distribution of the *Brca1* genotypes indicates the presence of several discrete genetic groups within the species *M. musculus*, which were formed under conditions of long-term isolation and accumulated during this time a series of mutations characteristic for each of them. However, later these groups widely radiated (perhaps with the involvement of humans), and began to hybridize in the contact zones.

Group I genotypes were found primarily in the house mice in the populations from Eastern Europe and Northern Trans-Urals; i.e., they are inherent of the subspecies *M. m. musculus sensu stricto*. Group III includes the genotypes of mice from Italy, undoubtedly belonging to the subspecies *M. m. domesticus*, the individuals of lines C57BL/6 and C57BL/6j, also originating from the last subspecies, as well as the mice from Armenia. Taking into account the information about the presence of a broad hybrid zone between *M. m. domesticus* and *M. m. musculus* in Transcaucasia [12, 37–41], the association of the genotypes of mice of C57BL/6 and C57BL/6j lines and mice from Italy and Armenia into one cluster is not surprising. Obviously, the genotypes of group III, despite the heterogeneity of the latter, have common roots and are characteristic of the subspecies *M. m. domesticus*. The genotypes constituting cluster II were found primarily in the mice from Central Asia (subspecies *M. m. wagneri*)

and Southern Siberia, including Transbaikalia, where animals nos. 424, 425, 437, and 447, similar in external attributes to *M. m. gansuensis*, were caught. Thus, it can be assumed that group II genotypes are characteristic of these two subspecies. The fact that group II is more closely related to group III (with *M. m. domesticus*) rather than with group I (*M. m. musculus*) contradicts the idea of the synonymy of *M. m. wagneri*, *M. m. gansuensis*, and *M. m. musculus* [3].

Mice nos. 1140 and 1214, which respectively formed separate basal branches IV and V of the dendrogram, originated from India, which is a part of the distribution area of the subspecies *M. m. castaneus*. However, the assumption that both mentioned individuals belong to this subspecies is questionable, because the genetic distance between them is relatively large ($D = 0.004$, see Table 3), and branches IV and V originate from different nodes of the tree. The BLAST analysis showed that the *Cox1* gene of mouse no. 1140 is less similar (98.5%) to the mtDNA sequences of members of the subspecies *M. m. castaneus* deposited in GenBank (KF781663 and KF781664) than the *Cox1* gene of individual no. 1214 (99.29%). Thus, specimen no. 1214 can be preliminarily regarded as belonging to the specified subspecies, whereas mouse no. 1140 is seemingly a representative of a certain unidentified subspecies or a genetic form. Northern India, where a high ancestral polymorphism of populations of synanthropic house mice was found, is considered their homeland [5, 38]. Our data confirm this assumption.

Hybrid individuals were found mostly in the contact zones of subspecies and contain the corresponding haplotypes. For example, the hybrids of *M. m. wagneri* and *M. m. musculus* were found in several localities in the Caspian region and in the southern regions of Western Siberia, and the hybrids of *M. m. wagneri* and *M. m. domesticus* as well as *M. m. musculus* and *M. m. domesticus* were found in Transcaucasia, i.e., just where they could be expected (Fig. 2). Several cases of identification of individuals (mostly hybrid) with a haplotype uncharacteristic of the populations of some territory can be explained by transfer of animals by transport or by the invasion of laboratory mice into natural populations, judging by the identity of one of the haplotypes of the hybrid specimen no. 16 from Ishim and the *Brca1* gene variant in several individuals of C57BL/6 and C57BL/6j lines (Fig. 1).

As already mentioned, the hypothetically determined unique haplotype of the hybrid specimen no. 447 from Transbaikalia (branch I-A) carries four substitutions that were not detected in any other studied mouse. Apparently, these mutations appeared in the recent past, because they are specific only to the animals of a certain individual genetic form. On the other hand, the same haplotype includes three additional nucleotides that are characteristic of the more ancient groups III–V but are missing in the *Brca1* gene

variants comprising groups I and II, adjacent and closest to branch I-A. This feature makes it possible to characterize the unique haplotype of mouse no. 447 as “pra-musculus”. It cannot be ruled out that it is a characteristic, though not absolute, of the subspecies *M. m. gansuensis* and will subsequently be found in some of its other members. The affiliation of the unique haplotype of hybrid individual no. 26925 (branch III-A), which carries five substitutions that were not found in any other mouse, to any subspecies or a genetically discrete form of *M. musculus* can be determined in two ways—either as one of the deviating variants of the *Brca1* gene within the subspecies *M. m. domesticus* (taking into account its high genetic heterogeneity) or as a haplotype that characterizes some other subspecies close to *M. m. domesticus* (e.g., *M. m. bactrianus*). To resolve this dilemma, the study of additional material from South Asia is required.

The greatest diversity of *Brca1* genotypes is characteristic of the populations of house mice from the Caspian region, which is consistent with the hypothesis of its proximity to the center of origin of the genus *Mus* and its location in the pathway of ancient migrations of the predecessors of modern subspecies [38].

The 5'-end region of the mitochondrial *Cox1* gene (approximately 700 bp from its start) is considered a good species-specific marker and is actively used in “DNA certification” of species and in phylogenetic reconstructions [42]. In the study of a close group of mice (wood mice of the genus *Sylvaemus*), the *Cox1* gene polymorphism analysis allowed successful discrimination between not only the species but also genetically allied intraspecific groups [43]. However, even the sufficiently extended *Cox1* fragment constituting more than three quarters of the entire gene appeared to be unsuitable for the identification of individuals of the subspecies *M. m. musculus*, *M. m. wagneri*, *M. m. gansuensis*, and *M. m. domesticus* from Transcaucasia, despite the fact that the animals were “pure” for the nuclear *Brca1* gene. Thus, our results of *Cox1* gene analysis are consistent with the data of other authors who studied the mitochondrial genes *D-loop* and *Cytb* [7, 16–18] and can be explained either by the stabilizing selection of mtDNA in house mice of the subspecies *M. m. musculus*, *M. m. wagneri*, and *M. m. gansuensis* or by large-scale introgression from any one subspecies to the gene pool of other subspecies due to hybridization [16]. It should be noted that the presence of foreign mtDNA can be traced in our sample in some individuals with well distinguishable *Cox1* gene mitotypes and best explains the differences in the topology of the dendrograms constructed on the basis of the nuclear and mitochondrial genes. For example, specimen no. 24383, a “pure” representative of the subspecies *M. m. wagneri* for the *Brca1* gene (group II), has the *Cox1* mitotype that is very close to the mitotype of mouse no. 1214 and the subspecies *M. m. castaneus*. A similar *Cox1* mitotype was found in specimen no. 26925,

which showed a high similarity with the individuals of the subspecies *M. m. domesticus* for nDNA.

Thus, the results of our study showed that the nuclear gene *Brca1* is a very promising molecular genetic marker to study variation, differentiation, and hybridization of house mice as well as to determine their subspecies affiliation. The *Cox1* gene, which evolves more rapidly than the *Brca1* gene, was found to be unsuitable to discriminate between individuals of *M. m. musculus*, *M. m. wagneri*, and *M. m. gansuensis* as well as the Transcaucasian representatives of *M. m. domesticus* due to introgression and long-term maintenance of foreign mtDNA in populations. However, *Cox1* gene analysis (along with diagnosis of animals by nDNA) may be useful in the evaluation of population differences within the subspecies *M. m. castaneus* and *M. m. domesticus*.

ACKNOWLEDGMENTS

We are sincerely grateful to O.V. Brandler (Koltzov Institute of Developmental Biology, Russian Academy of Sciences), A.Yu. Puzachenko (Institute of Geography, Russian Academy of Sciences), S.B. Pole (Kazakh Scientific Center for Quarantine and Zoonotic Diseases), A.A. Vlasov (Central Chernozem Reserve), E.I. Zholnerovskaya (Institute of Systematics and Ecology of Animals, Siberian Branch, Russian Academy of Sciences), V.I. Ermolaev (Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences), V.G. Shtro (Station of the Institute of Plant and Animal Ecology, Ural Branch, Russian Academy of Sciences, Labytnangi), and A.S. Saidov (Institute of Zoology and Parasitology, Academy of Sciences of Tajikistan) for their assistance in catching mice and laboratory processing of the material. We are also grateful to D.M. Shchepetov and V.G. Tambovtseva (Koltzov Institute of Developmental Biology, Russian Academy of Sciences) for their help with the molecular genetic analysis. Some of the animals were kindly provided by V.S. Lukarevskii and T.A. Mironova (Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences), A.Yu. Levykh (Ishim State Pedagogical Institute), Yu.A. Bazhenov (Dauriskii Reserve), A.A. Bannikova (Faculty of Biology, Moscow State University), A.E. Zykov (Kyiv), N.E. Burko, and I.S. Sadovnichaya (Minsk).

FUNDING

This study was performed within the State assignments of the Koltzov Institute of Developmental Biology, Russian Academy of Sciences (no. 0108-2019-0007, “Molecular Genetic and Environmental Mechanisms of Speciation and the Early Stages of Evolution. Development of Approaches to Estimating the Homeostasis of Development of Biological Systems (Methodology of Population Developmental Biology)”), Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences (no. 0109-2018-0073, “Evolutionary Morphology and Microevolution of Vertebrates”), and the Zoological Institute (no. AAAA-A17-117022810195-3,

“Phylogeny, Morphology, and Systematics of Placental Mammals”).

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interests. The authors declare no conflict of interest.

Statement on the welfare of animals. All procedures performed with the involvement of animals complied with the ethical standards of institutions or adopted practice of such studies.

REFERENCES

- Bonhomme F., Catalan J., Britton-Davidian J., Chapman V.M., Moriwaki K., Nevo E., Thaler L. 1984. Biochemical diversity and evolution in the genus *Mus*. *Biochem. Genet.* **22**, 275–303.
- Boursot P., Auffray J.-C., Britton-Davidian J., Bonhomme F. 1993. The evolution of house mice. *Annu. Rev. Ecol. Syst.* **24**, 119–152.
- Musser G.G., Carleton M.D. 2005. Superfamily Muridae. In *Mammal Species of the World: A Taxonomic and Geographic Reference*, 3rd ed. Eds Wilson D.E., Reeder D.M. Baltimore: Johns Hopkins Univ. Press, pp. 894–1531.
- Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature.* **420**, 520–562.
- Din W., Anand R., Boursot P., Darviche D., Dad B., Jouvin-Marche E., Orth A., Talwar G.P., Cazenave P.-A., Bonhomme F. 1996. Origin and radiation of the house mouse: Clues from nuclear genes. *J. Evol. Biol.* **9**, 519–539.
- Guénet J.-L., Bonhomme F. 2003. Wild mice: An ever-increasing contribution to a popular mammalian model. *Trends Genet.* **19**, 24–31.
- Suzuki H., Nunome M., Kinoshita G., Aplin K.P., Vogel P., Kryukov A.P., Jin M.-L., Han S.-H., Maryanto I., Tsuchiya K., Ikeda H., Shiroishi T., Yonekawa H., Moriwaki K. 2013. Evolutionary and dispersal history of Eurasian house mice *Mus musculus* clarified by more extensive geographic sampling of mitochondrial DNA. *Heredity.* **111**, 375–390.
- Hamid H.Sh., Darvish J., Rastegar-Pouyani E., Mahmoudi A. 2017. Subspecies differentiation of the house mouse *Mus musculus* Linnaeus, 1758 in the center and east of the Iranian plateau and Afghanistan. *Mammalia.* **81**, 147–168.
- Marshall J.T., Jr., Sage R.D. 1981. Taxonomy of the house mouse. *Symp. Zool. Soc. London.* **47**, 15–25.
- Sage R.D., Atchley W.R., Capanna E. 1993. House mice as models in systematic biology. *Syst. Biol.* **42**, 523–561.
- Marshall J.T., Jr. 1998. *Identification and Scientific Names of Eurasian House Mice and Their European Allies, Subgenus Mus (Rodentia: Muridae)*. Springfield, Virginia: National Museum of Natural History, pp. 1–80.
- Prager E.M., Orrego C., Sage R.D. 1998. Genetic variation and phylogeography of Central Asian and other house mice, including a major new mitochondrial lineage in Yemen. *Genetics.* **150**, 835–861.
- Geraldes A., Basset P., Gibson B., Smith K.L., Harr B., Yu H.-T., Bulatova N., Ziv Y., Nachman M.W. 2008. Inferring the history of speciation in house mice from

- autosomal, X-linked, Y-linked and mitochondrial genes. *Mol. Ecol.* **17**, 5349–5363.
14. Spiridonova L.N., Chelomina G.N., Moriwaki K., Yonekawa H., Bogdanov A.S. 2004. Genetic and taxonomic diversity of the house mouse *Mus musculus* from the Asian part of the former Soviet Union. *Russ. J. Genet.* **40** (10), 1134–1143.
 15. Yakimenko L.V., Korobitsyna K.V., Frisman L.V., Moriwaki K., Yonekawa H. 2003. Cytogenetics and systematics of mouse mice from Russia and neighboring countries. In *Problemy evolyutsii* (Problems of Evolution). Vladivostok: Dal'nauka, vol. 5, pp. 62–89.
 16. Spiridonova L.N. 2014. Introgression of nuclear and mitochondrial DNA markers of *Mus musculus musculus* to aboriginal populations of wild mice from Central Asia (*M. m. wagneri*) and South Siberia (*M. m. gansuensis*). *Mol. Biol.* (Moscow). **48** (1), 75–83.
 17. Suzuki H., Yakimenko L.V., Usuda D., Frisman L.V. 2015. Tracing the eastward dispersal of the house mouse, *Mus musculus*. *Genes Environ.* **37**, 20.
 18. Maltsev A.N., Stakheev V.V., Bogdanov A.S., Fomina E.S., Kotenkova E.V. 2015. Phylogenetic relationships of intraspecific forms of the house mouse *Mus musculus*: Analysis of variability of the control region (D-loop) of mitochondrial DNA. *Dokl. Biol. Sci.* **465**, 285–288.
 19. Brown W.M., George M., Wilson A.C. 1979. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 1967–1971.
 20. Wilson A.C., Cann R.L., Carr S.M., George M., Gyllenstein U.B., Helm-Bychowski K.M., Higuchi R.G., Palumbi S.R., Prager E.M., Sage R.D., Stoneking M. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.* **26**, 375–400.
 21. Bannikova A.A. 2004. Molecular markers and modern mammalian phylogeny. *Zh. Obshch. Biol.* **65**, 278–305.
 22. Lundrigan B.L., Jansa S.A., Tucker P.K. 2002. Phylogenetic relationships in the genus *Mus*, based on paternally, maternally, and biparentally inherited characters. *Syst. Biol.* **51**, 410–431.
 23. Suzuki H., Shimada T., Terashima M., Tsuchiya K., Aplin K. 2004. Temporal, spatial, and ecological modes of evolution of Eurasian *Mus* based on mitochondrial and nuclear gene sequences. *Mol. Phyl. Evol.* **33**, 626–646.
 24. Tucker P.K., Sandstedt S.A., Lundrigan B.L. 2005. Phylogenetic relationships in the subgenus *Mus* (genus *Mus*, family Muridae, subfamily Murinae): Examining gene trees and species trees. *Biol. J. Linn. Soc.* **84**, 653–662.
 25. Abel K.J., Xu J., Yin G.-Y., Lyons R.H., Meisler M.H., Weber B.L. 1995. Mouse *Brca1*: Localization sequence analysis and identification of evolutionarily conserved domains. *Hum. Mol. Genet.* **4**, 2265–2273.
 26. Sharan S.K., Wims M., Bradley A. 1995. Murine *Brca1*: Sequence and significance for human missense mutations. *Hum. Mol. Genet.* **4**, 2275–2278.
 27. Mammalian Gene Collection (MGC) Program Team. 2002. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16899–16903.
 28. Juvekar A., Hu H., Yadegarynia S., Lyssiotis C.A., Ullas S., Lien E.C., Bellinger G., Son J., Hok R.C., Seth P., Daly M.B., Kim B., Scully R., Asara J.M., Cantley L.C., Wulf G.M. 2016. Phosphoinositide 3-kinase inhibitors induce DNA damage through nucleoside depletion. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E4338–E4347.
 29. Yang H., Bell T.A., Churchill G.A., de Villena F.P.-M. 2007. On the subspecific origin of the laboratory mouse. *Nat. Genet.* **39**, 1100–1107.
 30. Yang H., Wang J.R., Didion J.P., Buus R.J., Bell T.A., Welsh C.E., Bonhomme F., Yu A.H.-T., Nachman M.W., Pialek J., Tucker P., Boursot P., McMillan L., Churchill G.A., de Villena F.P.-M. 2011. Subspecific origin and haplotype diversity in the laboratory mouse. *Nat. Genet.* **43**, 648–655.
 31. Sambrook J., Fritsch E.F., Maniatis T. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Lab. Press, pp. 1–398.
 32. Adkins R.M., Gelke E.L., Rowe D., Honeycutt R.L. 2001. Molecular phylogeny and divergence time estimates for major rodent groups: Evidence from multiple genes. *Mol. Biol. Evol.* **18**, 777–791.
 33. Nakamura I., Ohnuma A., Ichihashi T. 2008. Genbank Database. <http://www.ncbi.nlm.nih.gov/nuccore/193248428>.
 34. Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729.
 35. Bannikova A.A., Sighazeva A.M., Malikov V.G., Golenishchev F.N., Dzuev R.I. 2013. Genetic diversity of *Chionomys* genus (Mammalia, Arvicolinae) and comparative phylogeography of snow voles. *Russ. J. Genet.* **49** (5), 561–575.
 36. Rusin M., Lebedev V., Matrosova V., Zemlemerova E., Lopatina N., Bannikova A. 2018. Hidden diversity in the Caucasian mountains: An example of birch mice (Rodentia, Sminthidae, *Sicista*). *Hystrix*. **29**, 61–66.
 37. Mezhzherin S.V., Kotenkova E.V., Mikhailenko A.G. 1994. Hybrid zones. In *Domovaya mysh'. Proiskhozhdenie, rasprostranenie, sistematika, povedenie* (The House Mouse: Origin, Distribution, Systematics, Behavior). Moscow: Nauka, pp. 37–50.
 38. Boursot P., Din W., Anand R., Darvich D., Dad B., Von Deimling F., Talwar G.P., Bonhomme F. 1996. Origin and radiation of the house mouse: Mitochondrial DNA phylogeny. *J. Evol. Biol.* **9**, 391–415.
 39. Mezhzherin S.V., Kotenkova E.V., Mikhailenko A.G. 1998. The house mice, *Mus musculus* s.l., hybrid zone of Transcaucasus. *Z. Zaugetierk.* **63**, 154–168.
 40. Kotenkova E.V. 2002. Hybridization between synanthropic house mouse species and its evolutionary role. *Usp. Sovrem. Biol.* **122**, 580–593.
 41. Milishnikov A.N., Lavrenchenko L.A., Lebedev V.S. 2004. Origin of the house mice (superspecies complex *Mus musculus* sensu lato) from the Transcaucasia region: A new look at dispersal routes and evolution. *Russ. J. Genet.* **40** (9), 1011–1026.
 42. Ratnasingham S., Hebert P.D.N. 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Mol. Ecol. Notes.* **7**, 355–364.
 43. Bogdanov A.S., Stakheev V.V., Zykov A.E., Yakimenko V.V., Mal'kova M.G. 2012. Genetic variation and differentiation of wood mice from the genus *Sylvaemus* inferred from sequencing of the cytochrome oxidase subunit I gene fragment. *Russ. J. Genet.* **48** (2), 186–198.

Translated by M. Batrukova