ISSN 0026-8933, Molecular Biology, 2010, Vol. 44, No. 5, pp. 699–708. © Pleiades Publishing, Inc., 2010.
Original Russian Text © G.N. Chelomina, D.M. Atopkin, 2010, published in Molekulyarnaya Biologiya, 2010, Vol. 44, No. 5, pp. 792–803.

GENOMICS AND TRANSCRIPTOMICS

UDC 575.85:599.323.4

Molecular Genetic Evidence of a Deep Phylogenetic Discontinuity between the Asian and European Races of Pygmy Wood Mouse Based on the Mitochondrial Cytochrome *b* Gene Variation

G. N. Chelomina and D. M. Atopkin

Institute of Biology and Soil Sciences, Far East Division, Russian Academy of Sciences, Vladivostok, 690022 Russia; e-mail: chelomina@ibss.dvo.ru

Received December 2, 2009
Accepted for publication February 18, 2010

Abstract—The variation of mitochondrial DNA (mtDNA) cytochrome *b* gene sequences of pygmy wood mouse *Sylvaemus uralensis* (Pallas, 1811) from local populations of European Russia, West Siberia, and neighboring countries (Moldova, Kazakhstan, Uzbekistan, and Turkmenistan) has been studied. Phylogenetic analysis based on our results and GenBank data revealed two clusters of haplotypes: the Western clade, which was reliably subdivided into two sequence groups, and the Eastern clade with no significant differentiation. The clusters corresponded exactly to the European and Asian races of pygmy wood mouse recognized previously on the basis of biochemical and karyotypic variation. We suppose that the Asian race can be considered as an independent allopatric species. This concept is supported by the following evidence: the high interrace divergence level suggesting more than 1 Ma of divergent evolution, the absence of common haplotypes and the hiatus between the main peaks of the mismatch distribution, difference in the codon usage frequencies, fixed nucleotide substitutions in *cyt b*, as well as different amino acid sequences of cytochrome *b*. Only specimens of the western phylogenetic lineage should be classified as *S. uralensis* (Pallas, 1811), while *S. tokmak*, according to its first description (Severtsov, 1873), may be considered as the species uniting specimens of the eastern phylogenetic lineage.

DOI: 10.1134/S0026893310050043

Key words: mitochondrial DNA, cytochrome *b*, molecular evolution, allopatric speciation

INTRODUCTION

Investigation of genetic diversity of populations and evolutionary relationships among species is a topical issue in modern molecular biology and genetics. In recent years, the body of experimental data on primary structure of nuclear and mitochondrial DNA (mtDNA) has been rapidly growing, as well as the interest to species phylogeography. A number of European species of small mammals associated with temperate deciduous forests show phylogeographic differentiation and molecular evidence of demographic expansion [1–3]. This pattern of genetic variation is indicative of vicariant separation in different forest refugiums in glacial periods with a subsequent demographic expansion during postglacial colonization. The degree of phylogeographic differentiation varies among species, primarily depending on the mtDNA nucleotide substitution rate, as well as on the species capability to dispersion and characteristics of their habitats [4, 5].

Phylogeographic studies of two West Palearctic wood mouse species of the genus *Sylvaemus*—*S. sylvaticus*, and *S. flavicollis*, which are sympatric to *S. uralensis* in the European part of its area, showed that the genogeographic structure of the populations was strongly affected by Pleistocene global climate changes in West and East Europe. The different degree of phylogeographic discontinuity in these closely related species with unlike ecological preferences suggests they used different regional refugiums and different survival strategies during the Quaternary glacial periods [6]. In this study, we have investigated the geographical pattern of genetic variation of pygmy wood mouse *S. uralensis*, the only species of the *Sylvaemus* genus whose area includes vast forest territories both in the European and in the Asian part of the continent, varying in paleoclimatic history, landscape and biotope characteristics. According to our preliminary results, the two population complexes of *S. uralensis* representing the European and the Asian race differ considerably in their phylogeographic structure and demographic history and show a high level of phylogenetic segregation reaching the taxonomic significance on the mtDNA level [7, 8].

The objective of this study was to investigate the intraspecies genetic structure and to estimate the degree of the phylogenetic discontinuity between the two geographic races of pygmy wood mouse. Investigation of the species evolution over vast geographic areas also contributes to comparative phylogeography

and improves the understanding of allopatric speciation and intraspecies systematics. The study also involved data obtained from GenBank.

EXPERIMENTAL

Specimens. The study involved 63 specimens of pygmy wood mouse *Sylvaemus uralensis*, from both our and Dr. A.S. Bogdanov's personal collection, gathered during field research in different regions of Russia and Kazakhstan, as well as in Turkmenistan, Uzbekistan, and Moldova. In addition, mtDNA *cyt b* sequences of pygmy wood mouse were retrieved from the GenBank database. Altogether, 100 *cyt b* fragments (380 bp) and 34 full-size sequences (1140 bp; Table 1, 9–14) of pygmy wood mouse from 33 and 23 Eurasian locations, respectively, were studied. Genomic DNA was isolated from ethanol-fixed tissues. Amplification and sequencing of *cyt b* was performed with universal primers [15]. Sequencing was performed using a BigDye Terminator Cycle Sequencing Kit v. 3.1 on an automated laser ABI PRISM 310 sequencer.

Genetic distances between individual DNA sequences were calculated based on the number of nucleotide substitutions per position in pairwise comparison. Average nucleotide divergences within phylogeographic groups, the net intergroup divergence with intraspecies divergence correction, and their standard deviations were determined according to Nei and Kumar [16]. The estimates of the divergence time and the evolutionary age of mtDNA lineages were based on the divergence rate of 2.85% per 1 Ma, based on comparison of net divergence levels of mtDNA *cyt b* sequences and the age of wood mouse fossils [17].

Phylogenetic trees were constructed using the Neighbor-Joining (NJ) [18], UPGMA (Unweighted Pair Group with Arithmetic Averages), and Maximum Parsimony (MP) [19] algorithms with MEGA 3 software package [20]; the NJ model was chosen using the Akaike information criterion (AIC) [21] with the Modeltest v. 3.06 software [22].

The mtDNA diversity and the degree of divergence were calculated using the DnaSP software package [23]. The degree of interpopulation differentiation was assessed with the F_{ST} coefficient using the molecular variation analysis, AMOVA, with the ARLEQUIN 3.1 software [24].

The molecular clock hypothesis was tested using the Tajima estimates [25] for the evolution rates in paired DNA sequences respective to the outgroup sequences (MEGA 3 software package). The outgroups were the brown rat (*Rattus norvegicus*), the Gaidner's shrew-mouse (*Mus pahari*), the broad-toothed field mouse (*Apodemus mystacinus*), the Korean field mouse (*A. peninsulae*), and the striped field mouse (*A. agrarius*), Acc. nos. EU349782, AB09839, AF159394, AY389003, and AY389011, respectively. The obtained *cyt b* sequences of pygmy wood mouse were deposited in the GenBank database: complete sequences (1140 bp), Acc. nos. FN430738–430770; partial sequences (420 bp), FN433601–433643 and FN564438-39.

RESULTS AND DISCUSSION

In 63 specimens of pygmy wood mouse *S. uralensis* from different regions of European Russia and neighboring countries, we determined nucleotide sequences of a 420 bp long mtDNA fragment (positions 1–420 of the full-size murine *cyt b* sequence). In combination with GenBank data, the total sample (from 33 localities) included 100 nucleotide sequences 380 bp long (positions 13–392) that showed 49 *cyt b* haplotypes ($Hd = 0.959 \pm 0.00009$) due to mtDNA polymorphism ($Pi = 0.02931 \pm 0.00141$). In 33 specimens of pygmy wood mouse from 23 localities, we determined the complete *cyt b* sequence (1140 bp). In combination with GenBank data, this sample with a high degree of nucleotide variation ($Pi = 0.03609 \pm 0.00219$) comprised 34 sequences, each of them presenting a unique haplotype ($Hd = 1.0 \pm 0.00005$), mainly due to singleton nucleotide substitutions. The majority of mutations, both among partial (380 bp) and complete (1140 bp) sequences were transitions ($R = 3.595$ for either dataset) (Fig. 1).

The portion of each nucleotide in *cyt b* sequences of pygmy wood mouse declined as A (32.3%) ⟶ T (28.9%) ⟶ C (26.5%) ⟶ G (12.4%), and different codon positions were characterized with different nucleotide content. The first position contained all bases in approximately equal proportions, the second was thymine-rich (42.4%), and the third was rich in adenine (45.7%) and poor in guanine (2.5%). The nucleotide composition of *cyt b* gene of *S. uralensis* was similar to those of other mammals, including different wood and field mouse species, characterized with a low guanine content (12.2–12.8%) and approximately equal portions of the other three bases (A, 28–30%; T, 29–31%; C, 26–31%) [26–28]. In other taxons, the patterns are different: for instance, avian *cyt b* sequences are rich in cytosine (average, 54%) and poor in thymine (average, 11.3%) [15].

Phylogenetic reconstructions were based both on partial and complete *cyt b* sequences. In 100 partial (380 bp) sequences studied (including 37 sequences from GenBank), there were 58 variable sites, including 41 sites informative in the parsimony analysis. In 34 complete sequences (including one from GenBank), there were 134 variable sites, 103 of them informative in the parsimony analysis.

The NJ tree constructed based on partial *cyt b* sequences (Fig. 2a) showed that the haplotypes of pygmy wood mouse clearly fall into two reciprocally monophyletic clades with a high level of statistical support. The same topology was preserved when other algorithms (UPGMA, ML, MP, and ME), were used, indicating that the result was highly significant (data

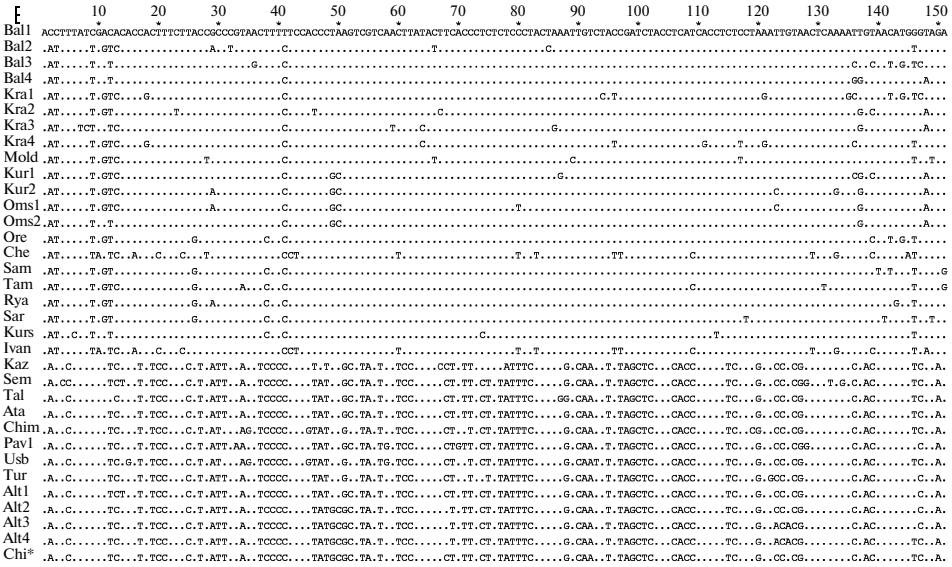

Fig. 1. Variable positions in the mtDNA *cyt b* sequence in *S. uralensis*. Dots indicate the identical positions. Notation as in Table 1.

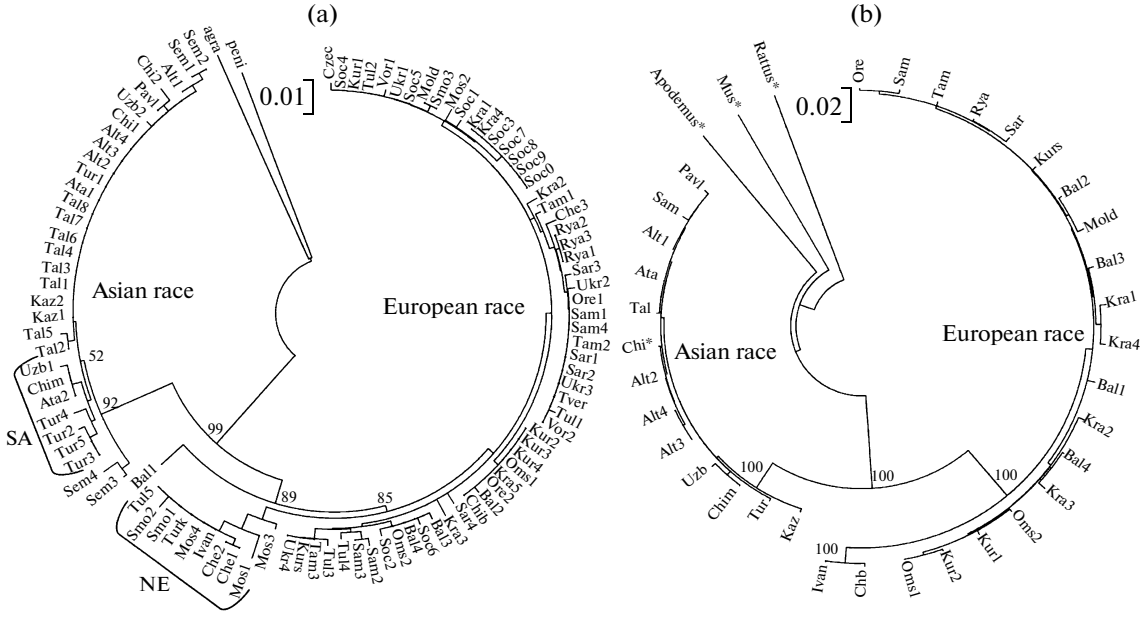

Fig. 2. Phylogenetic relationships among the *cyt b* haplotypes of *S. uralensis* based on the (a) partial (380 bp) and (b) complete (1140 bp) sequences constructed by NJ technique; bootstrap support values are given at nodes. SA is Southern haplotypes of the Asian race and NE is Northern haplotypes of the European race.

not shown). All haplotypes from Altai, Kazakhstan, Uzbekistan, Turkmenistan, and China (localities 1–10) form the Eastern clade; the other haplotypes representing the Russian plain, South Urals, West Siberia, Northern Caucasus, Moldova, the Czech Republic, and Turkey (localities 11–33) belong to the Western clade. The interclade nucleotide divergence *d* (with an intrarace divergence correction) is 3.8%. Within the Eastern clade, there was no phylogenetic division detected, although the haplotypes from the most Southern localities (4, 5, 7, and 8) formed a separate phylogenetic branch, although with insufficient statistical support. The Western clade included a highly significant subgroup comprising the haplotypes from most northern Russian localities (Chuvashia, Smolensk, Ivanovo, and Moscow oblasts; localities 12, 29, 18, and 24) and Turkey (locality 25). The phylogenetic clades detected correspond to the Asian and the Euro-

Table 1. Specimens, their origin, and genetic diversity

No.	Location	*n*	Code name	Complete sequence	Partial sequence	*Pi* (SD)	*Hd* (SD)
	Asian race ($n = 31$)		A			0.0064 (0.0011)	0.798 (0.074)
1	Kurchum region, Kazakhstan	2	Kaz	FN430738	FN433601	–	
2	Semipalatinsk oblast, Kazakhstan	4	Sem	FN430739	FN433602-FN433604	0.0109 (0.0029)	0.833 (0.222)
3	Taldy–Kurgan oblast, Kazakhstan	8	Tal	FN430740	FN433605-FN433611	0.0018 (0.0008)	0.464 (0.200)
4	Alma-Ata oblast, Kazakhstan	2	Ata	FN430741	FN433612	–	–
5	Chimkent oblast, Kazakhstan	1	Chim	FN430742		–	–
6	Pavlodar oblast, Kazakhstan	1	Pavl	FN430743		–	–
7	Angren neighborhood, Uzbekistan	2	Uzb	FN430744	FN433613	–	–
8	Airibaba neighborhood, Turkmenistan	5	Tur	FN430745	FN433614-FN433617	0.0058 (0.0015)	0.900 (0.161)
9	Gorno-Altaisk neighborhood, Altai	4	Alt	FN430746–FN430749		0.0013 (0.0007)	0.500 (0.265)
10	Uigur district, China	2	Chi	AY389021 [9]	AY532799 [data not published]	–	–
	European race ($n = 69$)		E			0.0129 (0.0010)	0.954 (0.011)
11	Orenburg oblast	2	Ore	FN430763	FN433620	–	–
12	Cheboksary neighborhood, Chuvashia	3	Che	FN430764	FN433618–FN433619	0.0175 (0.0083)	0.667 (0.314)
13	Samara oblast	4	Sam	FN430765	FN433621–FN433623	0.0048 (0.0014)	0.833 (0.222)
14	Tambov oblast	3	Tam	FN430766	FN433624–FN433625	0.0088 (0.029)	1.000 (0.272)
15	Ryazan' oblast	3	Rya	FN430767	FN433626–FN433627	0.0000 (0.000)	0.000 (0.000)
16	Saratov oblast	4	Sar	FN430768	FN433628; DQ844682-DQ844683 [11]	0.0105 (0.0048)	0.833 (0.222)
17	Kursk oblast	1	Kurs	FN430769		–	–
18	Ivanovo oblast	1	Ivan	FN430770		–	–
19	Nalchik neighborhood, Kabardino-Balkaria	4	Bal	FN430754–FN430757		0.0131 (0.0039)	1.000 (0.177)
20	Krasnodar krai	5	Kra	FN430758–FN430761	DQ844670 [11]	0.0079 (0.0018)	0.900 (0.161)
21	Omsk oblast	2	Oms	FN430750–FN430751		–	–
22	Kurgan oblast	4	Kur	FN430752–FN430753	FN564438–FN564439	0.0013 (0.0007)	0.500 (0.265)
23	Chisinau neighborhood, Moldova	1	Mol	FN430762		–	–
24	Moscow oblast	4	Mos	AF159393 [10]	DQ844678–DQ844680 [11]	0.0184 (0.0038)	1.000 (0.177)
25	Turkey	1	Turk	–	AJ311155 [12]	–	–
26	Chelyabinsk neighborhood	1	Chlb	–	AF127541 [13]	–	–
27	Khladovka, Ukraine	4	Ukr	–	EF016782–EF016785 [14]	0.0083 (0.0019)	1.000 (0.177)
28	Moravia, Czech Republic	1	Czec	–	AJ311154 [12]	–	–
29	Smolensk oblast	3	Smo	–	DQ844684–DQ844686 [11]	0.0175 (0.0068)	1.000 (0.272)
30	Tver oblast	1	Tver	–	DQ844681 [11]	–	–
31	Tula oblast	5	Tul	–	DQ844673–DQ844677 [11]	0.0153 (0.0036)	1.000 (0.126)
32	Voronezh oblast	2	Vor	–	DQ844671–DQ844672 [11]	–	–
33	Adygea and Krasnodar krai	10	Soch	–	DQ844660–DQ844669 [11]	0.00487 (0.0034)	0.756 (0.130)

Note: *n* is sample size and *Pi* is nucleotide diversity; columns 5 and 6 contain GenBank Acc. numbers of partial (380 bp) and complete (1140 bp) *cyt b* sequences; *Hd* is haplotype diversity and SD is standard deviation.

Table 2. Differences in the cytochrome *b* amino acid sequences between the geographic races of *S. uralensis*

Code name	23	39	**42**	136	156	**236**	327	329	**334**	356	359	**364**	366	367	369	371	372	373	380
Bal1	Ala	Ile	Thr	Val	Met	Ile	Ile	Val	Ile	Ile	Phe	Val	Met	Pro	Ser	Met	Ile	Glu	Asn
Bal2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bal3	.	.	.	.	.	.	.	.	.	.	.	.	.	Ser	Ala	Ile	.	.	.
Bal4	.	.	.	.	.	.	.	.	.	.	Cys	.	.	.	.	.	Asn	.	.
Kra1	.	.	Ala	.	.	.	.	.	.	.	.	.	.	Ser	Ala	Ile	.	.	.
Kra2	.	.	.	.	.	.	.	.	.	.	Cys	.	.	.	.	.	Asn	.	.
Kra3	.	.	.	.	.	.	.	.	.	.	Cys	.	.	.	.	.	Asn	.	.
Kra4	.	.	Ala	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mold	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Asp	.
Kur1	.	.	.	Gly	.	.	.	.	.	.	Cys	.	.	.	.	.	Asn	.	.
Kur2	.	.	.	Gly	.	.	.	.	.	.	Cys	.	.	.	.	.	Asn	.	.
Oms1	.	.	.	Gly	.	.	.	.	.	.	Cys	.	.	.	.	.	Asn	.	.
Oms2	.	.	.	.	.	.	.	.	.	.	Cys	.	.	.	.	.	Asn	.	.
Ore	.	.	.	.	.	.	.	.	.	.	.	.	.	Ser	Ala	.	.	.	.
Che	Thr	Met	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sam	.	.	.	.	.	.	.	.	.	.	.	.	Leu	Ser	.	.	.	.	Ser
Tam	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Ser
Rya	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sar	.	.	.	.	.	.	.	.	.	.	.	.	Ile	.	.	.	.	Asp	.
Kurs	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Ivan	Thr	Met	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Asn	.	.
Kaz	.	.	Met	.	Ile	Val	.	.	Val	.	.	Ile	.	.	.	Ile	.	.	.
Sem	.	.	Met	.	Ile	Val	.	.	Val	Val	.	Ile	.	.	.	Ile	.	.	.
Tal	.	.	Met	.	Ile	Val	.	.	Val	Val	.	Ile	.	.	.	Ile	.	.	.
Ata	.	.	Met	.	Ile	Val	.	.	Val	.	.	Ile	.	.	.	Ile	.	.	.
Chim	.	.	Met	.	.	Val	.	.	Val	.	.	Ile	.	.	.	Ile	.	.	.
Pav1	.	.	Met	.	Ile	Val	.	.	Val	.	.	Ile	.	.	.	Ile	.	.	.
Uzb	.	.	Met	.	.	Val	.	.	Val	.	.	Ile	.	.	.	Ile	.	.	.
Tur	.	.	Met	.	.	Val	.	.	Val	.	.	Ile	.	.	.	Ile	.	.	.
Alt1	.	.	Met	.	Ile	Val	.	.	Val	.	.	Ile	.	.	.	Ile	.	.	.
Alt2	.	.	Met	Gly	Ile	Val	.	.	Val	.	.	Ile	.	.	.	Ile	.	.	.
Alt3	.	.	Met	Gly	Ile	Val	Met	Ile	Val	.	.	Ile	.	.	.	Ile	.	.	.
Alt4	.	.	Met	Gly	Ile	Val	Met	Ile	Val	.	.	Ile	.	.	.	Ile	.	.	.
Chi*	.	.	Met	Gly	Ile	Val	.	.	Val	.	.	Ile	.	.	.	Ile	.	.	.

Note: sites with fixed substitutions are given in bold.

* GenBank database. Code names are as in Table 1.

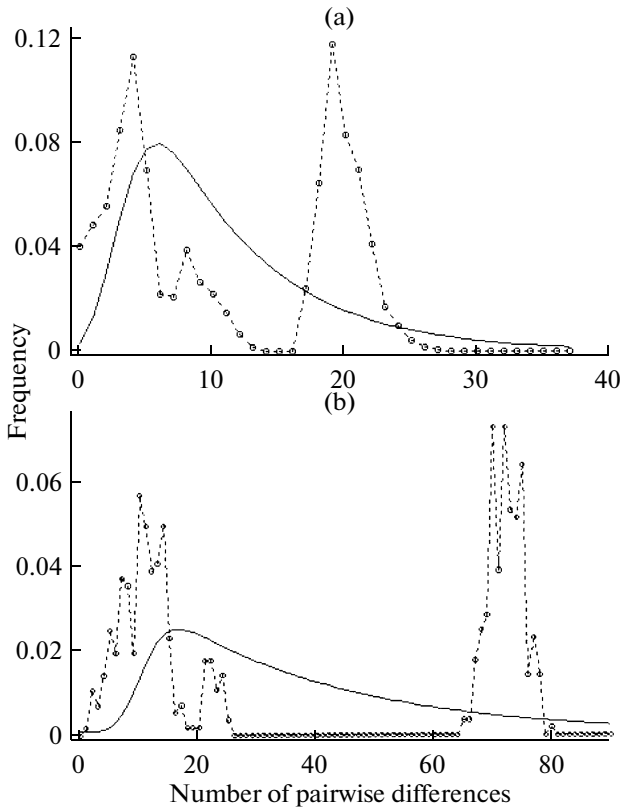

Fig. 3. Pairwise genetic difference distribution in the (a) partial (380 bp) and (b) complete (1140 bp) *cyt b* gene sequences of *S. uralensis*. Solid line is expected and dotted line is observed.

pean races of pygmy wood mouse described previously based on the karyologic and biochemical data [29–30].

Phylogenetic analysis based on complete *cyt b* sequences (using different algorithms) confirmed a strong differentiation between the geographic races of pygmy wood mouse (Fig. 2b). The genetic distance between the two population complexes (with intrarace divergence correction) was 5.8%. The interspecies divergence of murine *cyt b* sequences varies from 6 to 15% [27, 31]. The interspecies divergence values of the cytochrome oxidase I (COI) gene in different animal groups lie within a similar range (6–23%) [32]. Thus, the genetic difference between the races of pygmy wood mouse approaches the lower limit of interspecies difference in murine species. Exact correlation between the genetic distances and the conventional species limits according to the genetic species concept was confirmed in different rodent genera [33]. Theoretically, different estimates for phylogeographic divergence obtained with two datasets may result from different sample sizes or from an uneven distribution of mutations in *cyt b*.

The distribution of nucleotide substitutions in all pairs of *S. uralensis* haplotypes indicated a strong variation among mtDNA sequences representing different phylogeographic groups (Fig. 3). The hiatus observed between the more closely related and the more diverged mtDNA sequences suggests a high degree of genetic discontinuity among the population groups compared. Bimodal distribution of pairwise nucleotide difference is typical for samples including two spatially isolated populations [34].

The existence of two highly genetically diverged regional groups of *S. uralensis* is supported by a comprehensive analysis of the nucleotide and amino acid substitution patterns and the genetic diversity distribution between the races. In the European race, the nucleotide diversity of both partial and complete *cyt b* sequences ($Pi_{380} = 0.01285 \pm 0.00098$; $Pi_{1140} = 0.01161 \pm 0.00154$) was much higher than in the Asian race ($Pi_{380} = 0.00636 \pm 0.00107$; $Pi_{1140} = 0.00652 \pm 0.0021$). The haplotype diversity of partial *cyt b* sequences was fairly high, and was less pronounced when the races were considered individually ($Hd = 0.798 \pm 0.071$ in the Asian race and $Hd = 0.954 \pm 0.011$ in the European race). No such difference was detected in complete sequence comparison (all sequences were unique), probably, because nearly all localities were represented by a single mtDNA specimen. Since differences in genetic diversity can be related to differences in sample sizes, we calculated the respective parameters for each local sample of size 3 and more (Table 1). Local populations had a fairly high level of mtDNA haplotype diversity varying from 0.5 to 1; nucleotide diversity varied from 0.013 to 0.0184. The only exception was the Ryazan sample (GenBank data) characterized with a single haplotype and, consequently, the zero levels of haplotype and nucleotide diversity. The highest levels of nucleotide diversity were observed in the northernmost European localities: in Chuvashia, Smolensk, and Moscow oblasts. The average levels of both nucleotide and haplotype diversity of *cyt b* sequences were significantly higher among local populations of the European race ($Pi = 0.00987$, $Hd = 0.806$) than among Asian populations ($Pi = 0.00495$, $Hd = 0.675$).

Remarkably, there was not a single haplotype common between the two geographic races, suggesting that the sorting of mtDNA lineages has been completed. In addition, the races differed significantly in the haplotype frequency distribution (based on the partial sequence data). One of the 15 Asian haplotypes was found in half of the specimens from different localities, while the other haplotypes were found in isolated cases. The European haplotypes fall into three groups of different frequencies. Most of them (24 of 34) were low-frequency haplotypes found in a small number of cases, three haplotypes were found in approximately one third of specimens (i.e., they have the highest frequency), while the remaining seven accounted for approximately one fourth of all animals, showing intermediate frequencies.

Comparison of nucleotide and gene diversity parameters can give an insight into the demographic

history of populations and allow speculations concerning past demographic events [34]. For instance, high values of both parameters (as observed in the European race) reflect a large population size and a wide area. A high genetic and low nucleotide diversity (which is a more appropriate description for the Asian race) can result from a long isolation of a relatively small population. Low values of both parameters suggest that the population has descended from an efficiently small ancestor population, that had experienced a period of depression [34–35].

An analysis of the genetic diversity distribution based on partial *cyt b* sequences using an exact test for population differentiation and AMOVA, involving haplotype frequencies and interhaplotype nucleotide diversity, confirmed the conclusion about the strong differentiation between the geographic regions (races): Fst = 0.797, $P < 0.001$, 79.72% of the total variance. Similar results were obtained with complete *cyt b* sequences: Fst = 0.845, $P < 0.001$, 79.82% of the total variance. Respectively, for both datasets, the values of the gene flow between the phylogeographic groups determined from nucleotide diversity were essentially small ($N_m < 1$), which is a further evidence of genetic discontinuity between the geographic groups of mtDNA haplotypes in *S. uralensis*.

In all field and wood mouse species studied, *cyt b* coding sequence begins with a conserved methionine-encoding ATG codon [27]. A comparative analysis of codon usage (based on the complete sequence data) showed that the European race used 50 and the Asian race used 48 codons of the 60 mitochondrial amino acid-encoding codons. Nine codons are not used by either race (CGG/Arg, CGC/Arg, UCG/Ser, AGU/Ser, ACG/Thr, CAG/Gln, GCG/Ala, GAG/Glu, and GUG/Val), and four codons are not used by one of the races (CCG/Pro, AAG/Lys, and UUG/Trp by the Asian, and UGU/Cys by the European). Only one third of all codons are used with equal frequencies. The most frequent codons are the same in both races: CUA (Leu), AUC (Ile), UUC (Phe), and AUU (Ile), but their actual frequencies differ and are 31, 27, 18, 16% (wide range) in the Asian race and 27, 20, 21, 22% (narrow range) in the European race, respectively. Both races prefer mitochondrial codon variant for Met and Trp.

Interestingly, in both races *cyt b* position 163 is occupied by Thr, but it is encoded with a universal codon in the European race, and with a mitochondrial codon (UGA) in the Asian race; in the nuclear genome it encodes the termination of polypeptide synthesis. It was recently shown that, in the standard genetic code, UGA can also be interpreted in more than one way: a stop codon by default, it can be translated into the 21st amino acid, selenocysteine, if the coding region is followed by a selenocysteine insertion sequence, SECIS [36]. Codon usage bias, or codon preference, is a common phenomenon observed in a wide range of organisms, including prokaryotes, animals and plants, reflecting a balance of mutations, genetic drift, and the natural selection for translation optimization (efficiency and/or precision) [37]. Although mammals, yeast, and *E. coli* differ in codon preference, related species have similar preference types. Optimal codons are those most frequently used in genes with high expression levels. The usage of these codons reduces the ribosome binding time and/or the amino acid misincorporation rate in comparison to alternative synonymic codons. In general, the codon preference in different species may affect (via gene expression) the evolution rates in individual lineages [37, 38].

A comparison of amino acid sequences of cytochrome *b* (derived from complete nucleotide sequences) in European and Asian races of pygmy wood mouse revealed 19 phylogenetically relevant substitutions, four of which (mainly, between aliphatic amino acids Ile and Val) were fixed (Table 2). Notably, an analysis of a larger sample including 126 partial amino acid sequences confirmed the interrace divergence at position 42. Most of the amino acid sequences studied also showed interrace difference at positions 156 and 371. Interestingly, cytochrome *b* of *Apodemus mystacinus* (GenBank), the most evolutionary ancient species of West Palearctic wood mouse, is the only other protein to contain Ile at position 236 among all *Sylvaemus* species studied.

Some authors suggest that the *cyt b* sequences of *Apodemus* wood and field mouse species deposited in GenBank include cryptic pseudogenes [39]. Nuclear copies of mitochondrial genes ("numt"—nuclear mitochondrial pseudogenes) have been described in many animal and plant species; assumedly, such species are not infrequent in mammalian genomes [40, 41]. However, the absence of stop codons (including those that encode amino acids in nuclear genome), the preference for mitochondrial variants of amino acid codons, as well as the results of phylogenetic analysis (all sequences of *S. uralensis* form a single cluster separated from other related species, including sympatric ones) are strong arguments against the presence of pseudogenes in our sequence sample [8].

Assumedly, at the beginning of the last glacial maximum, the phylogeographic structure of the contemporary European mammalian species was little pronounced, and the currently recognized phylogeographic patterns are relics of the last glacial period and do not reflect the long-term environmental adaptation [42]. Therefore, an association between the phylogenetic structure and the geographic distribution does not necessarily suggest that there has been a long-term genetic isolation [43]. However, our results clearly indicate the existence of two allopatric groups that separated many thousand of years before the last glacial maximum. Assuming that mtDNA of wood mouse accumulates approximately 2.85% of nucleotide difference per 1 Ma [17], the observed nucleotide divergence between the two clades (based on the partial and

complete *cyt b* sequences) suggests that the separation occurred approximately 1.5 Ma. This timing corresponds to the Danube glacial period at the beginning of Pleistocene (1.5–1.8 Ma) [44].

We should point out several factors affecting the estimates of intrataxon phylogenetic discontinuity (see, e.g., [5, 6]), such as incompleteness of the paleontological record for a number of species, including wood and field mice, the importance of accounting for the whole range of the species area, and the representativeness of the samples studied. Phylogeographic histories of species with vast Eurasian areas were seldom investigated, partially because of the difficulties associated with collecting the material over large territories [35]. The timing depends on the uniformity of the molecular evolution rate, and the results obtained based on the standard molecular clock should be treated with caution [45].

Although the glacial period in East Eurasia was short, this region remained extremely cold and arid, the most of its territory being large tundra and steppe zones [44, 46]. The available data on global climate changes suggest that 2 Ma ago North China and North Kazakhstan were covered in loessial soils indicative of arid climate and predomination of short-grass steppe landscapes [47]. Since pygmy wood mouse is predominantly associated with humid forest biotops, the above factors might have formed the primary ecological barrier that divided the ancestor *S. uralensis* form into two geographic groups in the early Pleistocene. Secondary contacts during interglacial periods could have aggravated the interrace genetic divergence, as it has been proposed for many other West European animal species [48].

During the Quaternary climatic optimums, a single broad-leaved woodland belt existed in Eurasia at least twice: in Kazantsev (West Siberia), or Mikulin (East Europe), Pleistocene interglacial period, and in Atlantic and Subboreal Holocene periods; however, the Kazantsev period was the last time when the oak and the associated fauna had a continuous Transpalearctic area. Palynologic data suggest that, in the warmer Kazantsev period in South Siberia, the broad-leaved forests (oak, elm, lime) had a wide range spreading continuously beyond 60° North (cited as in [49]). Thus, in the Kazantsev period, migration between the pygmy wood mouse races was fairly possible; however, in the Holocene optimum it was hardly probable because of the degradation of the broad-leaved forests. The absence of West to East migration in the Holocene climatic optimum of many Eurasian species can be explained by the earlier expansion optimum of broad-leaved woodland in East Eurasia [49].

Phylogenetic discontinuity in the south-east part of the area and evidence of demographic expansion that reflect isolation accompanied with colonization from two different refugium sources have been observed in many Eurasian species, and in some cases the molecular divergence among populations reaches the level of taxonomic significance. Among the boreal forest species, the deepest phylogeographic discontinuity (10.5%) was found in Siberian salamander, *Salamandrella keyserlingii*; the most insignificant was found in wood lemming *Myopus schistocolor* (0.9%) and in Siberian flying squirrel *Pteromys volans* (0.7%); great tit *Parus major* (5.6%) and great spotted woodpecker *Dendrocopos major* (3.0%) showed intermediate values (cited as in [5]).

The phylogenetic structure of related wood and field mouse species (*Sylvaemus*), studied at present, is different. In European wood mouse *S. sylvaticus*, common in Central, West, and East Europe, the subdivision is well pronounced. The segregation between its two phylogenetic mtDNA clades (one haplotype lineage is restricted to the Balkans; the other one is widespread) is dated to 1.5–1.6 Ma, based on genetic distances [6]. These data suggest that the geographic structure of genetic variation in *Sylvaemus* species with a wide European or Eurasian range was affected by common historical factors.

Thus, the phylogeography of the pygmy wood mouse apparently reflects the species' mitochondrial gene pool structure. We detected a deep phylogeographic discontinuity suggesting that two lineages of pygmy wood mouse were separated in different refugiums for the first Pleistocene glacial periods. The genetic divergence between the mtDNA clades probably involves the polymorphism that had existed in the ancestor population since the early Pleistocene. Apparently, allopatric division of mtDNA haplotypes was preserved because of the gene flow barriers and the low dispersion rate, and was supported by maternal philopatry and populational isolation, as supposed, for instance, for the brown bear [50] and the Eurasian badger [51]. It appears that the western and eastern haplotype groups were isolated from gene flow and began to diverge; it was the isolation that was responsible for the genetic hiatus between the races and complete mtDNA lineage sorting, indicating the end of the initial stage of allopatric speciation. Fixed nucleotide substitutions altering the amino acid sequence of cytochrome *b* are another evidence of deep phylogenetic discontinuity between the races. Our data on mtDNA *cyt b* sequence variation in pygmy wood mouse, together with the earlier results [8], suggest that the interrace differentiation has reached the speciation level, although the results of RAPD analysis speak rather in favor of subspeciation [7]. This pattern of genetic diversity distribution corresponds to allopatric speciation in the Pleistocene [52] and is supported by the biostratigraphic data [49].

Thus, our data suggest that the geographic races of the pygmy wood mouse can be considered as individual taxons of the species level. Moreover, only specimens of the western phylogenetic clade can be classified as *S. uralensis* (Pallas, 1811), with a single gene pool acknowledged for the whole range from the Czech Republic to South Urals (its typical territory).

Population complexes of the eastern phylogenetic clade (and probably the *palliles* [8] form) apparently can be united as the species named *S. tokmak* (Severtsov, 1873). It is this form of pygmy wood mouse that was first described in the Asian part of its range (North Kyrgyzstan). The notion of these taxons being young allopatric species is further substantiated by their polytypism. In addition, young allopatric species often have bordering areas [53]. In general, *cyt b* can be used as a reliable molecular marker for further systematic analysis of pygmy wood mouse forms and their phylogenetic relationships.

ACKNOWLEDGMENTS

We are grateful to Dr. A.S. Bogdanov (Institute of Developmental Biology, Russian Academy of Sciences) for kindly providing specimens from the pygmy wood mouse tissue collection.

REFERENCES

1. Arbogast B.S., Kenagy G.J. 2001. Comparative phylogeography as an integrative approach to historical biogeography. *J. Biogeography*. **28**, 819–825.
2. Deffontaine V., Libois R., Kotlik P., Sommer R., Nieberding C., Paradis E., Searle J.B., Michaux R. 2005. Beyond the Mediterranean peninsulas: Evidence of central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). *Mol. Ecol*. **14**, 1727–1739.
3. Kotlik P., Deffontaine V., Mascheretti S., Zima J., Michaux J.R., Searl J.B. 2006. A northern glacial refugium for bank voles (*Clethrionomys glareolus*). *Proc. Natl. Acad. Sci. USA*. **103**, 4860–4864.
4. Goropashnaya A.V., Fedorov V.B., Seifert B., Pamilo P. 2004. Limited phylogeographic structure across Eurasia in two red wood ant species *Formica pratensis* and *F. lugubris* (Hymenoptera, Formicidae). *Mol. Ecol*. **13**, 247–276.
5. Fedorov V.B., Goropashnaya A.V., Boeskorov G.G., Cook A. 2008. Comparative phylogeography and demographic history of the wood lemming (*Myopus schistocolor*): Implications for Late Quaternary history of the taiga species in Eurasia. *Mol. Ecol*. **17**, 598–610.
6. Michaux J.R., Libois R., Filipucci M.G. 2005. So close and so different: Comparative phylogeography of two small mammal species, the yellow-necked fieldmouse (*Apodemus flavicollis*) and the woodmouse (*Apodemus sylvaticus*) in the Western Palaearctic region. *Heredity*. **94**, 52–63.
7. Atopkin D.M., Bogdanov A.S., Chelomina G.N. 2007. Genetic variation and differentiation in stripped field mouse *Apodemus agrarius* inferred from RAPD-PCR analysis. *Russ. J. Genet*. **43**, 665–676.
8. Chelomina G.N., Atopkin D.M., Bogdanov A.S. 2007. Phylogenetic relationships between species and intraspecific forms of forest mice from the genus *Sylvaemus* as determined by partial sequencing of the cytochrome *b* gene of mitochondrial DNA. *Doklady Biol. Sci*. **416**, 356–359.
9. Liu W., Wei F., Li M., Jiang X., Feng Z., Hu J. 2004. Molecular phylogeny and taxonomy of wood mice (genus *Apodemus* Kaup, 1829) based on complete mtDNA cytochrome *b* sequences, with emphasis on Chinese species. *Mol. Phylogenet. Evol*. **33**, 1–15.
10. Martin Y., Gerlach G., Meyer A. 2000. Molecular phylogeny of European muroid rodents based on complete cytochrome *b* sequences. *Mol. Phylogenet. Evol*. **16**, 37–47.
11. Balakirev A.E., Baskevich M.I., Gmyl' A.P., Okulova N.M., Andreeva N.A., Sokolenko O.V., Malygin V.M., Khlyap L.A., Oparin M.L., Orlov V.N. 2007. On the taxonomic rank of *ciscaucasicus* and its relationships with the pygmy wood mouse *Sylvaemus uralensis* inferred from the mtDNA cytochrome *b* gene sequence. *Genetika*. **43**, 1651–1666.
12. Michaux J.R., Chevret P., Filippucci M.G., Macholán M. 2002. Phylogeny of the genus *Apodemus* with a special emphasis to the genus *Sylvaemus* using the nuclear IRBP gene and two mitochondrial markers: Cytochrome *b* and 12S rRNA. *Mol. Phylogenet. Evol*. **23**, 123–136.
13. Makova K.D., Nekrutenko A., Baker R.J. 2000. Evolution of microsatellite alleles in four species of mice (genus *Apodemus*). *J. Mol. Evol*. **51**, 166–172.
14. Hoofer S.R., Gaschak D., Dunina-Barkovskaya Y.V., Makluk J., Meeks H.N., Wickliffe J.K., Baker R. S. 2007. New information for systematics, taxonomy, and phylogeography of the genus *Apodemus* (*Sylvaemus*) in Ukraine. *J. Mammology*. **88**, 330–342.
15. Kocher T.O., Thomas W.K., Meyer A. 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. USA*. **86**, 6196–6200.
16. Nei M., Kumar S. 2004. *Molecular Evolution and Phylogenetics*. NY: Oxford Univ. Press.
17. Michaux J.R., Magnanou E., Paradis E., Nieberding C., Libois R. 2003. Mitochondrial phylogeography of the woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. *Mol. Ecol*. **12**, 685–697.
18. Saitou N., Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. **4**, 406–425.
19. Nei M. 1987. *Molecular Evolutionary Genetics*. NY: Columbia Univ. Press.
20. Kumar S., Tamura K., Nei M. 2004. MEGA 3: An integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinformat*. **5**, 150–163.
21. Akaike H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr. AC-19*, 716–723.
22. Posada D., Crandall K.A. 1998. Modeltest: Testing the model of DNA substitution. *Bioinformatics*. **14**, 817–818.
23. Rozas J., Sanchez-DelBarrio J.C., Messequer X., Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*. **19**, 2496–2497.
24. Excoffier L., Laval G., Schneider S. 2006. *Arlequin ver. 3.1: An Interrated Software Package for Population*

Genetics Data Analysis. Switzerland: Inst. Zool. Comp. Mol. Pop. Gen. Lab. (CMPG).

25. Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. **123**, 585–595.

26. Irwin D.M., Kocher T.O., Wilson A.C. 1991. Evolution of the cytochrome *b* gene of mammals. *J. Mol. Evol*. **32**, 128–144.

27. Martin Y. Gerlach G., Schlotterer C., Meyer A. 2000. Molecular phylogeny of European muroid rodents based on complete cytochrome *b* sequences. *Mol. Phylogenet. Evol*. **16**, 37–47.

28. Chelomina G.N., Suzuki H. 2006. Molecular evolution and phylogeography of Northern Palearctic *Sylvaemus* mice as inferred from data on variability of nuclear (IRBP) and mtDNA cytochrome *b* gene. *Zool. Zh*. **85**, 219–234.

29. Bogdanov A.S. 2001. Chromosome differentiation of populations of the pygmy wood mouse, *Sylvaemus uralensis*, in the eastern part of the species range. *Zool. Zh*. **80**, 331–342.

30. Bogdanov A.S. 2004. Allozyme variation of pygmy wood mouse *Sylvaemus uralensis* (Rodentia, Muridae) and estimation of the divergence of its chromosome forms. *Russ. J. Genet*. **40**, 897–909).

31. Chelomina G.N. 2005. *Wood and field mice: Molecular-genetic aspects of evolution and systematics*. Vladivostok: Dal'nauka.

32. Shneyer V.S. 2007. On the specificity of DNA: Fifty years after. *Biochemistry*. **72**, 1377–1384.

33. Bradley R.D., Baker R.J. 2001. A test of the genetic species concept: Cytochrome *b* sequences and mammals. *J. Mammol*. **82**, 960–973.

34. Avise J. C. 2000. *Phylogeography: The History and Formation of Species*. Cambridge, MA: Harvard Univ. Press.

35. Painter J.N., Siitonen J., Hanski I. 2007. Phylogeographical patterns and genetic diversity in three species of Eurasian boreal forest beetles. *Biol. J. Linn. Soc*. **91**, 267–279.

36. Turanov A.A., Lobanov A.V., Fomenko D.E., Morrison H.G., Sogin M.L., Klobutcher L.A., Hatfield D.L., Gladyshev V.N. 2008. Genetic code supports targeted insertion of two amino acids by one codon. *Science*. **323**, 259–261.

37. Ingvarsson P.K. 2007. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*. *Mol. Biol. Evol*. **24**, 836–844.

38. Grantham R., Gautier C., Gouy M., Mercier R., Pavé A. 1980. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res*. **8**, 49–62.

39. Dubey S., Michaux J., Brünner H., Hutterer R., Vogel P. 2009. False phylogeny on wood mice due to cryptic cytochrome-*b* pseudogene. *Mol. Phylogenet. Evol*. **50**, 633–641.

40. Mirol P.M., Mascheretti S., Searle J.B. 2000. Multiple nuclear pseugenes of mitochondrial cytochrome *b* in *Ctenomys* (Caviomorpha, Rodentia) with either great similarity to or high divergence from the true mitochondrial sequence. *Heredity*. **84**, 538–547.

41. Traint D.A., DeWoody A.J. 2008. Molecular analyses of mitochondrial pseudogenes within the nuclear genome of arvicoline rodens. *Genetica*. **132**, 21–33.

42. Hofreiter M., Serre D., Rohland N., Rabeder G., Nagel D., Conard N., Münzel S., Pääbo S. 2004. Lack of phylogeography in European mammals before the last glaciation. *Proc. Natl. Acad. Sci. USA*. **35**, 12963–12968.

43. Leonard J.A., Wayne R.K., Cooper A. 2000. Population genetics of Ice Age brown bears. *Proc. Natl. Acad. Sci. USA*. **97**, 1651–1654.

44. Tarasov P.E., Volkova V.S., Webb T., et al. 2000. Last glacial maximum biomes reconstructed from pollen and plant macrofossil data from northern Eurasia. *J. Biogeography*. **27**, 609–620.

45. Arbogast B.S., Edwards S.V., Wakeley J., et al. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annu. Rev. Ecol. Syst*. **33**, 707–740.

46. Anderson B.C., Borns Jr.H.W. 1997. *The Ice Age World. An Introduction to Quaternary History and Research with Emphasis on North America and Europe during the Last 2.5 Million Years*. Oslo, Norway: Scand. Univ. Press.

47. Zubakov V.A. 1990. *Global'nye klimaticheskie sobytiya Neogena* (Global Climatic Events in the Neogene). Leningrad: Gidrometeoizdat.

48. Hewitt G.M. 2004. Genetic consequences of climatic oscillations in the Quatenary. *Phil. Trans. R. Soc. London. B*. **359**, 183–195.

49. Dubatolov V.V., Kosterin O.E. 2000. Nemoral species of Lepidoptera (Insecta) in Siberia: A novel view on their history and the timing of their range disjunctions. *Entomol. Fennica*. **11**, 141–166.

50. Barnes I., Matheus P., Shapiro B., Jensen D., Cooper A. 2002. Dynamics of Pleistocene population extinction in Beringian brown bears. *Science*. **295**, 2267–2270.

51. Marmi J., López-Giráldez F., Macdonald D.W., Calafell F., Zholnerovskaya E., Domingo-Roura X. 2006. Mitochondrial DNA reveals a strong phylogeographic structure in the badger across Eurasia. *Mol. Ecol*. **15**, 1007–1020.

52. Hare M. 2001. Prospect for nuclear gene phylogeography. *Treds Ecol. Evol*. **16**, 700–706.

53. Palumbi S.R., Lessios H.A. 2005. Evolutionary animation: How do molecular phylogenies compare to Mayr's reconstruction of speciation patterns in the sea? *Proc. Natl. Acad. Sci. USA*. **102**, 6566–6572.