



Research Article

Hidden diversity in the Caucasian mountains: an example of birch mice (Rodentia, Sminthidae, *Sicista*)

Mikhail RUSIN^{1,*}, Vladimir LEBEDEV², Vera MATROSOVA³, Elena ZEMLEMEROVA^{4,5}, Natalia LOPATINA⁶, Anna BANNIKOVA⁵

¹Schmalhausen Institute of Zoology, B. Khmel'nitskogo 15, 01030 Kiev, Ukraine

²Zoological Museum, Moscow State University, B. Nikitskaya 6, 125009 Moscow, Russia

³Engelhardt Institute of Molecular Biology, Vavilova 32, 119991 Moscow, Russia

⁴A.N. Severtsov Institute of Ecology and Evolution, Leninskij prosp. 33, 119071 Moscow, Russia

⁵Moscow State University, L. Gory, 119234 Moscow, Russia

⁶Institute of Systematics and Ecology of Animals SB RAS, Frunze st. 11, 630091 Novosibirsk, Russia

Keywords:

Sicista
Caucasus
cryptic species
phylogeny
molecular clock

Article history:

Received: 24 April 2017

Accepted: 7 August 2017

Acknowledgements

The research was supported by Mohamed bin Zayed Species Conservation Fund (projects 150510512 and 14258509 to MR), phylogenetic analysis was performed by VL within the framework of the Russian Science Foundation, project No.14-50-00029. Acknowledgements extend to A. Martynov, A. Tikhonov, N. Nedyalkov, T. Pushkar, P. Weinberg, J. Tekeev, T. Hayrapetyan and G. Papov for the help during field trips, A. Ivanova for laboratory assistance. Special thanks to Yu. Kovalskaya, A. Surov, I. Tikhonov, G. Ryurikov and M. Oparin for donating several samples of birch mice. Also we are grateful to anonymous reviewers for their critical reviews that helped to make our work much better.

Abstract

For the first time all members of *Sicista caucasica* species complex were genotyped using one mitochondrial and five nuclear markers. We revealed that there are two lineages in the group: western (*caucasica* + *kluchorica*) and south-eastern (*armenica* + *kazbegica*). This phylogeographic pattern corresponds with recent findings on several other species of small mammals of the Caucasus (moles, snow voles, dormice). An unexpectedly high hidden diversity is found within *kazbegica* and *kluchorica*. Also, we confirmed the presence of highly divergent cryptic species in *S. tianschanica*. The results of the molecular clock analysis suggest that the *S. caucasica* group separated from its sister taxon — the *S. betulina-subtilis* group — in the Late Pliocene. Major division within the *S. caucasica* group occurred in the Upper Early Pleistocene (~1.16 Mya). The split in the south-eastern lineage leading to modern *S. kazbegica* and *S. armenica* should have happened at ~1 Mya; the division of the western lineage took place in the Middle Pleistocene (~570 Kya). The approximate ages for the basal splits in *S. kazbegica* and *S. kluchorica* are estimated at 230 Kya and 350 Kya, respectively.

Introduction

Birch mice (genus *Sicista*) are one of the least known group of small mammals in Eurasia. This may be due to their secretive nature, low numbers and highly fragmented spatial distribution. Phylogenetically *Sicista* is placed as the basal branch in the superfamily Dipodoidea (Lebedev et al., 2013; Pisano et al., 2015). There are around 13–14 species recognized in the genus, although, the existence of yet undescribed species cannot be ruled out. The genus is currently distributed over Palearctic only. While some species have large distribution area (like *Sicista betulina*), others are limited to few localities. The latter category includes species of the Caucasian mountain region. Most of the birch mice show minor morphological differences between each other but, at the same time, they are often separated by significant chromosomal or genetic distances. Based on this a high level of cryptic diversity has been revealed in the genus. Originally all unstriped *Sicista* were included in the species *Sicista concolor*. Later Vinogradov (1925) described a separate species *Sicista caucasica*. In 1980th based on chromosomal variability a single species was divided into four. The nominative form *S. caucasica sensu stricto* has 2n=32, nFa=46, three other species were named as *S. kazbegica* with 2n=40–42, nFa=48–50, *S. armenica* 2n=36, nFa=50 and *S. kluchorica* 2n=24, nFa=42 (Sokolov et al., 1981, 1986; Sokolov and Baskevich, 1988, 1992). From that time on a species complex called *Sicista caucasica* group is recognized and

is believed to consist of four karyologically distinct but morphologically uniform species (Fig. 1). Noteworthy, due to very limited information on these species and small number of localities they are currently included to the Red List of Threatened Species of IUCN as Endangered (*armenica* and *kazbegica*), Vulnerable (*caucasica*) or Near Threatened (*kluchorica*).

After the discovery of this hidden diversity, the group was ignored in molecular phylogenetic studies for a long time. Finally, two publications of independent groups of researchers appeared in 2015 (Pisano et al., 2015; Baskevich et al., 2015, 2016). Result of the first work was based on single *S. caucasica* and *S. kluchorica* specimens with two specimens of *S. kazbegica* (representing as 2n=40, so 2n=42 chromosomal races) (Pisano et al., 2015). The authors of the second study used just slightly larger sampling: one specimen of *S. caucasica*, two specimens of *S. kazbegica* (although both represent 2n=40 chromosomal race) and three *S. kluchorica* (Baskevich et al., 2016). Unfortunately, *S. armenica* has not been covered by any researches so far. The origin of the samples used in the study of Pisano et al. (2015) was unclear as well. According to F. Catzeffis (pers. comm., 2016) these specimens were donated by M. Baskevich and some of them represent the same populations as used in her publication (Baskevich et al., 2016). While Baskevich et al. (2016) used a single cytochrome b mitochondrial DNA marker; Pisano et al. (2015) employed multilocus analysis: both cytochrome b and several nuclear DNA markers (IRBP, BRCA1, GHR, RAG1). From results of both these teams it was unveiled that *S. kluchorica* and *S. caucasica* form a pair of sister taxa, and *S. kazbegica*

*Corresponding author

Email address: mrusin@list.ru (Mikhail RUSIN)



Figure 1 – Four species of birch mice of the Caucasus: A - *Sicista armenica*; B - *S. kazbegica*; C - *S. kluchorica*; D - *S. caucasica*. All photos by MR.



Figure 2 – Geography of studied samples of *Sicista caucasica* group: *S. caucasica* (circles); *S. kluchorica* (filled squares denotes a genotyped locality, empty square - non-genotyped terra typica); *S. kazbegica* (diamond, Kazbek; terra typica); *S. armenica* (filled circle denotes a genotyped locality, empty circle - non-genotyped terra typica). Approximate extent of occurrence of each species is shown as dashed lines. Major geographic regions mentioned in the text: Ad - Republic of Adygea; Kr - Krasnodar Kray; KBR - Republic of Kabardino-Balkaria; NOAR - Republic of North Ossetia-Alania; ChR - Chechen Republic; Da - Dagestan; 1 - Arkhyz locality and Kizgych River; 2 - Bezengi Wall in Eastern Balkaria; 3 - Ardon River.

appeared to be the most remote species in the group. The phylogenetic position of *S. armenica* remained unknown.

Thus, our aim was to reconstruct the phylogeny of the *Sicista caucasica* group using both mitochondrial and nuclear DNA markers and to provide estimates of divergence times based on the molecular clock.

Materials and methods

Field sampling

DNA sampling of *S. caucasica* group was performed in 2014–2016 years in Greater Caucasus (in Russian Federation: Republics of North Ossetia–Alania, Karachay–Cherkessia and Adygea) and Lesser Caucasus (Armenia) (Fig. 2). All animals were captured using pitfall traps. Tissue samples (one digit per animal) were preserved in 95% ethanol. After sampling all animals were released back to nature. Deceased animals were taken into museum collection. Totally we collected 10 samples of all four species of *S. caucasica* group: *armenica* (1 specimen), *kazbegica* (2 specimen), *kluchorica* (4 specimen) and *caucasica* (3 specimen). Additionally, as comparative material we included samples of 13 specimens collected either by authors or donated by colleagues: *S. subtilis* (4), *S. lorigera* (3), *S. betulina* (2), *S. strandi* (3), *S. tianschanica* (1). Details of all specimens used in the study are given in Tab. S1. Thus totally we analyzed 23 specimens.

As Caucasian birch mice represent a group of sibling species in the identification of our animals we had to rely on known range information. Due to restrictions on working with Endangered animals we were

not able to perform a cytogenetic analyzes as it requires sacrificing live specimens.

DNA isolation, PCR amplification and sequencing

Genomic DNA from ethanol-preserved tissues was extracted using a standard protocol of proteinase K digestion, phenol–chloroform deproteinization and isopropanol precipitation (Sambrook et al., 1989). We sequenced the complete mitochondrial cytochrome b (cytb) gene and fragments of five nuclear loci: exon 11 of the breast cancer type 1 susceptibility protein (BRCA1), exon 1 of the interphotoreceptor binding protein gene (IRBP), intron 2 of the thyrotropin gene (THY), intron 13 of the betaspectrin 1 gene (SPTBN), intron 9 of the protein kinase C gene (PRKC). Nucleotide sequences of the original primers specially designed for amplification and sequencing are provided in the Tab. S2. The PCR protocol for all genes was initial denaturation at 94 °C for 3 min, then 30 cycles of 94 °C for 30 s, 52–65 °C (depending on the primer pair) for 1 min, and 72 °C for 1 min, with a final extension of 72 °C for 6 min. PCR products were visualized on 1.5% agarose gel and then purified using ammonium-ethanol precipitation. Approximately 10–30 ng of the purified PCR product was used for sequencing with each primer by the autosequencing system ABI 3100-Avant using the BigDye™ Terminator Chemistry v. 3.1 (Applied Biosystems, Foster City, CA, USA). Assembling was performed using SeqMan (Lasergene, USA). The sequences obtained in this study were deposited in the GenBank (accession numbers see in Tab. S1).

Alignment and partitioning

All sequences were aligned by eye using BioEdit v. 7.0.9.0 (Hall, 1999). Phylogenetic reconstructions were performed with the following data sets: cytb alignment containing 33 sequences; IRBP and BRCA1 alignments of 25 and 33 sequences respectively; nuclear concatenation containing sequences of three introns and two exons for 13 specimens; nuclear and mitochondrial cytb sequences combined in a species-tree estimation under multispecies coalescent model.

In the analyses employing nuclear concatenation (as well as in separate analyses of BRCA1 and IRBP alignments) all sequences were used as unphased genotypes with heterozygous position coded using the IUB ambiguity codes. For species tree reconstruction the genotype data on each of the five nuclear genes were phased using Phase software (Stephens et al., 2001) in combination with DNAsp ver.5 (Librado and Rozas, 2009).

The program PartitionFinder (Lanfear et al., 2012) was used to determine the optimum partitioning scheme for each protein-coding gene under BIC criterion. The best-fit partitioning scheme for the cytb suggests subdivision into three subsets corresponding to codon positions. The BRCA1 and IRBP subsets were partitioned into two subsets with 1st and 2nd codon positions combined.

Comparing our sequences with all available data from GenBank demonstrated that some of the sequences may contain errors. In particular, sequences of cytb (Acc.N. KR107025–KR107032) seem to have erroneous ends; therefore, we did not use in our study positions 1–34 and 1119–1140. Sequence Acc.N. KP715878 should be missing one nucleotide in polyA region at position 16–21.

Tree reconstruction and molecular dating

Maximum likelihood (ML) reconstructions were conducted in Treefinder, version October 2008 (Jobb 2008). Appropriate models of sequence evolution were selected for each of the subsets under Bayesian information criterion (BIC) employing the routine implemented in Treefinder. Clade stability was tested based on 1000 pseudoreplicates.

Bayesian tree reconstructions were performed in MrBayes 3.2 (Ronquist et al., 2012). Models with either two or six rate matrix parameters were selected for each subset based on the results of the model selection for the ML analysis. The analysis included two independent runs of four chains with the default heating scheme. The chain length was set at five million generations with sampling every 2000 generation. Tracer 1.6 software (Rambaut and Drummond, 2005) was used to check for

convergence and determine the necessary burn-in fraction, which was 10% of the chain length. The effective sample size exceeded 200 for all estimated parameters.

The species tree was reconstructed employing a Bayesian coalescent framework as implemented in *BEAST (Heled and Drummond, 2010). For either gene the hLRT tests performed in PAML ver 4.7 (Yang, 2007) did not reject the hypothesis of rate constancy. Therefore, the analysis in *BEAST was performed under strict clock. Partitioning and substitution models were as in the ML analysis. A Yule prior for the species tree shape and the piecewise constant population size model were assumed. Default priors were used for all other parameters. Two runs of 100 million generations were conducted. Parameter convergence was assessed in Tracer.

The tree was calibrated using the estimates of the substitution rates of BRCA1 and IRBP obtained in a previous phylogenetic analysis of Dipodidae: the prior density of substitution rates of IRBP and BRCA1 was modeled using gamma distribution with the mean and standard deviation equal to those of the posterior obtained in a previous study (BRCA1: mean 2.63e-3, st.dev 2.69e-4; IRBP: mean 3.39e-3, st.dev 4.41e-4) (Shenbrot et al., 2017).

We had to accept this approach because of the lack of reliable fossil calibrations. Although there are many birch-mice fossils of Pleistocene and Neogene age it remains unclear whether any of them can be attributed unambiguously to contemporary species or species groups. In particular, it seems inappropriate to use the age of *Sicista primus* Kimura, 2011 (~17 Mya) as a proxy for the time of the basal split among recent species of *Sicista* as it was done in Zhang et al. (2013). The cladistic analysis of dental characters Kimura (2013) recovered this species as a sister group of all other examined fossil and recent birch mice, albeit with just moderate support. This result suggests that the age of the split between *S. primus* and other taxa can be significantly older than the time of the most recent ancestor of contemporary *Sicista*.

Results

Genetic structure

Sequences of one mtDNA and five nDNA markers were obtained for up to 23 individuals belonging to nine species; final alignments included the original data and available sequences from GenBank. Totally from 17 to 33 individuals were included into genetic analysis of different markers (Tab. S3). *Cytb* appeared to be the most divergent of all studied markers (40% of positions are variable). All nuclear markers showed variability of 6–10%. Heterozygotes were found in all nuclear markers except THY.

Mitochondrial gene diversity

Putative pseudogene of cytb in Sicista betulina

The *cytb* sequences of both specimens of *S. betulina* that were initially obtained with the L7/H6 primer combination contained many ambiguities (positions with two clear peaks of approximately equal intensity). To separate the two products a set of specific primers (L300b, H458b) was designed and additional sequencing runs were performed. As a result, we recovered two sequences with the p-distance between them being 7.4%. Neither variant contained any stop codons or frame shifts. A comparison with the Genbank data showed that one of the paralogues (variant B) had no close match (>93%). The second paralogue (variant A) appeared very close (~99.7%) to the single Genbank sequence attributed to *S. betulina* (KP715861) and is also similar (~98.8–98.7%) to the published sequences of *S. strandi* (KP715862, KP715863). However, the inspection of the translated alignment showed that the position #263 contains a stop codon (AGG) in both KP715862 and KP715863. In all other *Sicista* this position corresponds to serin amino acid. Therefore, we believe that the paralogue A is, in fact, a *cytb* pseudogene, i.e. a nuclear copy of mitochondrial DNA also known as NUMT (Bensasson et al., 2001). This conclusion is supported further by the fact that the sequences of *S. strandi* obtained in this study are rather distant from both A and B clusters, thus, suggesting that the low level of divergence observed between the Genbank sequences of *S. strandi* and *S. betulina*

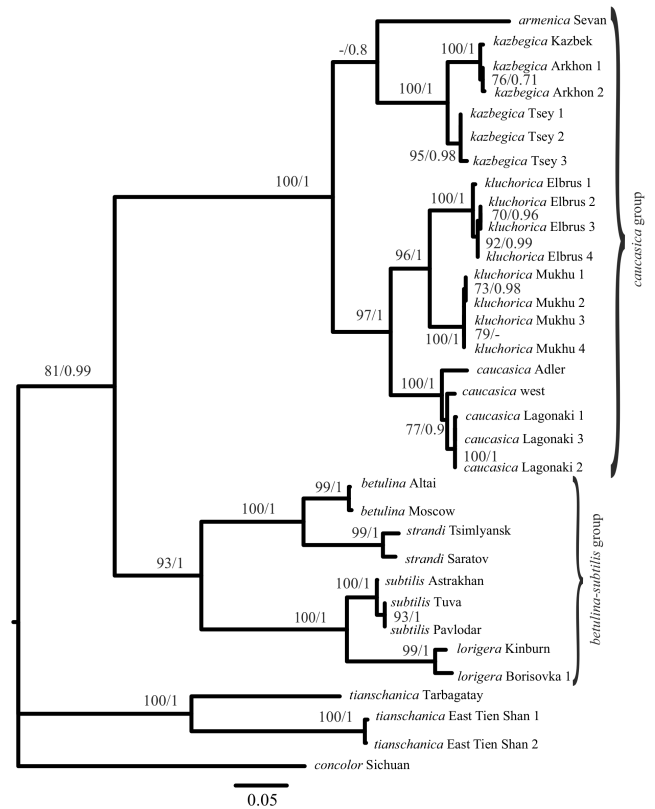


Figure 3 – The ML phylogenetic tree of *Sicista* reconstructed from the *cytb* data. Sample names correspond to Tab. S1. Values at nodes represent ML bootstrap support based on 1000 pseudoreplicates / Bayesian posterior probability.

may be a result of a slowdown of molecular evolution of a nuclear pseudogene. Based on these considerations we excluded all sequences of the cluster A from the phylogenetic analyses.

Phylogenetic reconstruction based on *cytb*

Maximum likelihood and Bayesian inference analyses yielded identical topologies (Fig. 3). The robustly supported Caucasian clade is recovered as the sister group of the *betulina*–*subtilis* group. *Sicista kluchorica* and *S. caucasica* cluster together with high support. *Sicista armenica* is placed sister to *S. kazbegica* but the support for this association is low (BS=57%, PP=0.80). *Sicista kluchorica* is found to contain two divergent sublineages originating from Mukhu and Elbrus localities, respectively, and separated by a genetic distance (K2p) of ~6% (Tab. 1). Likewise, *S. kazbegica* as well includes two groups showing ~5% pairwise divergence. However, the divergence between the two lineages of *S. tianschanica* is even larger (16%), which is comparable to that between *S. caucasica* and *S. kazbegica* or *S. subtilis* and *S. betulina*.

Nuclear gene diversity

The phylogenetic trees constructed from the alignments of BRCA1 and IRBP (Fig. S4) robustly support the monophyly of the caucasica group but show less resolution within it. Sister-group relationship between *caucasica* and *betulina*–*subtilis* groups is supported by IRBP but not BRCA1. Both genes demonstrate high level of divergence between the lineages of *S. tianschanica*. The concatenation (3814 bp) of five nuclear genes provides sufficient resolution to all nodes (Fig. 4). The topology agrees with that of the *cytb* tree. The position of *armenica* as sister to *kazbegica* is strongly supported.

Species tree and molecular clock

The species tree reconstructed by *BEAST algorithm (Fig. 5) recapitulates the topology inferred from nuclear concatenation. The age of the root node is estimated at ca. 4.4. Mya (95% HPD: 3.4–5.6 Mya; Early Pliocene). The date of the split between the Caucasian clade and

Table 1 – K2p distance matrix of cytb mitochondrial gene between sampled populations of different *Sicista* species.

<i>concolor</i>														
<i>tianschanica</i>	0.22													
<i>tarbagatai</i>	0.21	0.16												
<i>betulina</i>	0.21	0.22	0.20											
<i>strandii</i>	0.22	0.22	0.22	0.10										
<i>subtilis</i>	0.22	0.23	0.21	0.17	0.16									
<i>lorigera</i>	0.24	0.22	0.21	0.18	0.17	0.10								
<i>armenica</i>	0.23	0.23	0.25	0.21	0.23	0.25	0.25							
<i>kazbegica</i> tsey	0.22	0.24	0.26	0.21	0.21	0.23	0.25	0.14						
<i>kazbegica</i> kazbek+arkhon	0.22	0.24	0.26	0.21	0.22	0.23	0.24	0.15	0.05					
<i>kluchorica</i> elbrus	0.22	0.22	0.23	0.19	0.19	0.20	0.22	0.16	0.14	0.15				
<i>kluchorica</i> mukhu	0.22	0.23	0.23	0.19	0.19	0.20	0.21	0.17	0.15	0.15	0.06			
<i>caucasica</i> lagonaki	0.23	0.24	0.23	0.20	0.18	0.21	0.21	0.16	0.14	0.14	0.10	0.09		
<i>caucasica</i> west	0.24	0.24	0.24	0.20	0.19	0.20	0.21	0.15	0.14	0.14	0.10	0.10	0.01	
<i>caucasica</i> adler	0.24	0.24	0.24	0.20	0.19	0.21	0.22	0.17	0.15	0.15	0.10	0.10	0.03	0.03

the *betulina*–*subtilis* group is estimated as ca. 3.6 Mya (2.7–4.6 Mya). *Sicista betulina* and *S. subtilis* branches diverge at ca. 1.8 Mya (1.35–2.4 Mya; Early Pleistocene). The age of the basal split in the Caucasian clade is estimated as 1.16 Mya (0.85–1.5 Mya; late Early Pleistocene). The split between *armenica* and *kazbegica* is just slightly younger – ca. 1.0 Mya (0.71–1.3 Mya), while the separation of *kluchorica* from *caucasica* dates back to Middle Pleistocene – ca. 570 Kya (390–780 Kya).

The *BEAST estimated the substitution rate for the cytb gene as 0.11 per site per My (95% HPD: 0.079–0.14). Based on this value one can tentatively estimate the time of divergence of the two branches of *S. tianschanica* as ca. 1.3 Mya. The approximate ages of the basal splits in *kazbegica* and *kluchorica* are 230 Kya and 350 Kya, respectively. Since the rate estimate may be biased due to rate decay (Ho et al., 2005) the dates for the recent events may be overestimated.

Discussion

Phylogeny of *Sicista caucasica* group

In the present study all *Sicista* species known from the Caucasus were covered for the first time. The phylogenetic position of *caucasica* group as obtained from our multilocus analyzes is concordant with findings of Pisano et al. (2015). Inside the group two large lineages are identifiable. The first one is *armenica* + *kazbegica* (south-eastern clade) and the second is *kluchorica* + *caucasica* (western clade). The monophyly of the second group has been demonstrated already (Pisano et al., 2015; Baskevich et al., 2016). Concerning *armenica* there have been no molecular data present so far, while neither karyology nor morphology could clearly reveal the position of the species within the group.

Cytogenetics

The inferred phylogeny of the *caucasica* group appears to be concordant with known cytogenetic data. The available mitochondrial data support the subdivision of *kazbegica* into two lineages, which most probably correspond to two variants of its karyotype: 2n=42 NFa=50 from the Kazbek (terra typica) and 2n=40 NFa=48 from Tsey Gorge (Sokolov and Baskevich, 1992). Sister group relationship between *caucasica* and *kluchorica*, which was demonstrated also in previous molecular studies (Pisano et al., 2015; Baskevich et al., 2016), is in agreement with the similarity of their karyotypes established from chromosome banding data (Baskevich et al., 2004). Concerning the Armenian birch mouse the latter authors believed that, from the cytogenetic perspective, it occupies “intermediate position” between *kazbegica* and *caucasica* + *kluchorica*. The karyotype of *armenica* (2n=36 NFa=50) shares the same fundamental number with the more primitive karyotype variant of *kazbegica*. Hypothetically, the difference between the two karyotypes can be explained by three centric fusions and a pericentric inversion. In contrast, *kluchorica* is separated from *kazbegica* by nine non-centric fusions and two inversions (Baskevich et al., 2004). By supporting the

monophyly of *armenica* + *kazbegica* the molecular data suggests that shared karyotype characters such as common NFa may be regarded as potential synapomorphies.

Cryptic diversity within species

We were able to identify a previously unknown phylogeographic complexity of *S. kluchorica*. Previously this species was supposed to be uniform from Arkhyz Reserve to Elbrus Mountain with type locality (Kluchor Pass at the right bank of Teberda River) being in between these two points. We studied samples from Mukhu River locality (left bank of Teberda River) and compared them to specimens from Elbrus. The K2p distance between them is high enough (6%) and may even correspond to a value characteristic for sibling species in Rodentia (Baker and Bradley, 2006). Comparable distance was one of the main reasons to declare *S. trizona* and *S. lorigera* as separate species (Cserkés et al., 2016). Thus, it may be tempting to declare specimens from Mukhu as representatives of a nominative *kluchorica* form based on a small geographic distance between Mukhu and Kluchor localities and treat populations from Elbrus as a new species or a subspecies. But it would be incorrect at this stage of knowledge. Indeed, Mukhu and Kluchor localities are located close to each other (less than 30 km), but on the different sides of Teberda River. Examples of Ardon and Kizgych rivers show that somehow they can present major barriers for dispersal of birch-mice lineages. Thus, any taxonomic conclusions are currently premature. The IRBP data adds even more confusion to relationship between Mukhu and Elbrus populations: despite limited sampling at least one animal from Mukhu is closer to Elbrus sample.

Two chromosomal races have been already known (Sokolov and Baskevich, 1992) in *kazbegica*. We can assume that these races may correspond to two distinct genetic sublineages. It was suggested that the 40-chromosomal race from Tsey Gorge lives to the North of the main ridge of the Great Caucasus in Russia (Republics of North Ossetia and Chechnya) and the 42-chromosomal race lives to the South of the Great Caucasus in Georgia (Kazbek locality) (Baskevich, n.d.). Our data do not support this hypothesis. Although we do not have karyotype data for our studied samples from Arkhon Pass (North Ossetia, Russia), the molecular markers clearly show that they are closely related to specimens from terra typica in Georgia and not to those from Tsey Gorge in North Ossetia. We can suggest that *kazbegica* shares the same phylogeographic pattern of other *Sicista* species in the region. All of the forms replace each other in direction from east to west: *kazbegica* Kazbek – *kazbegica* Tsey – *kluchorica* Elbrus – *kluchorica* Mukhu – *caucasica*. The only exception to this “rule” is the geographic position of *armenica* as it occupies the Lesser Caucasus. Preliminary, the border between sublineages of *kazbegica* lies in the deep valley of Ardon River. Such hypothesis may correspond with findings of Baskevich et al. (2004) that Kizgych River forms a barrier between *kluchorica* and *caucasica*. *Sicista caucasica* appears to be the least variable among

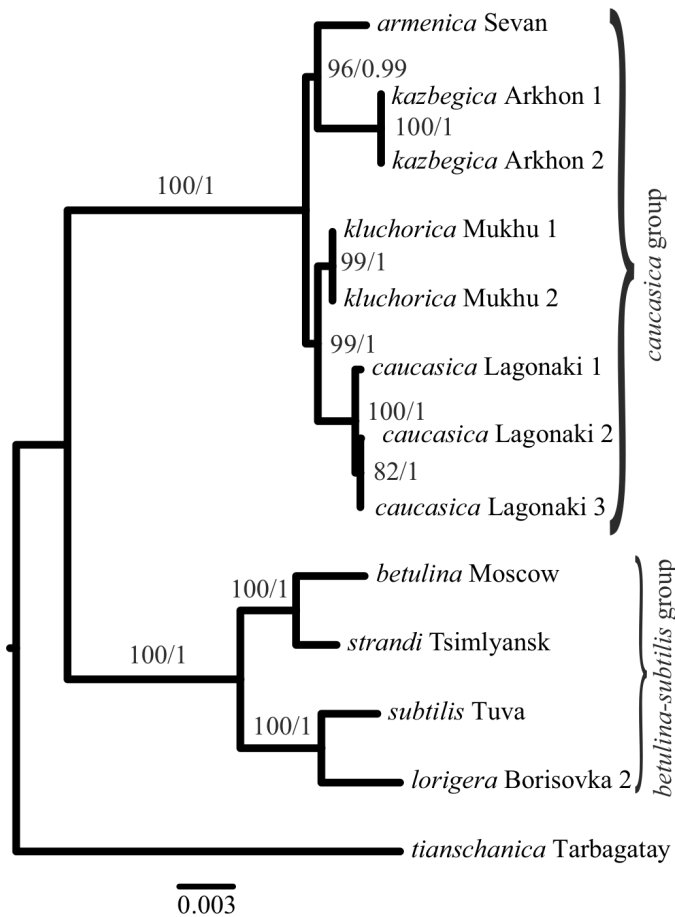


Figure 4 – The ML phylogenetic tree of *Sicista* based on the concatenated alignment of five nuclear loci (384 bp). Sample names correspond to Tab. S1. Values at nodes represent ML bootstrap support based on 1000 pseudoreplicates / Bayesian posterior probability.

all Caucasian birch mice; with an obvious exception of *S. armenica* as there is a single animal genotyped up to date.

Comparative phylogeography

The phylogeographic pattern observed in the *S. caucasica* group may be tracked down in several other taxa of small mammals. All of them show a separation into two major clades – western and eastern. The precise position of the boundary between them is often yet unknown (as in *Sicista* case) but is located typically in Central Caucasus. For example, western subspecies of Levant moles *Talpa levantis* complex — *T. l. minima* — is described from Adygea and has around 3% distance from a lineage comprising moles from Kabardino-Balkaria and Armenia (Bannikova et al., 2015). In this case, the border between western and eastern lineages lies to the west of the Elbrus Mountain. Even more interesting example is presented by snow voles *Chionomys gud* and *C. roberti*. Both species show phylogeographic structure similar to that of the *S. caucasica* group: the major dichotomy separates eastern clade (Eastern Balkaria, North Ossetia, Dagestan and Georgia) from the western clade, which consists of two well supported subclades (samples from Elbrus and Adygea, respectively). The border between western and eastern lineages of *C. gud* is located between the Elbrus Mountain and Bezengi Wall in Eastern Balkaria (Bannikova et al., 2013). Another example is illustrated by the forest dormouse *Dryomys nitedula*, which also shows the presence of two lineages: the western one comprising of animals from Adygea, Krasnodar and Arkhyz and the Central Caucasian one occurring in Elbrus and adjacent region (Grigoryeva et al., 2015). It is possible that there could be more species with a high cryptic diversity in the Caucasus that has the same phylogeographic pattern as in the *S. caucasica* group.

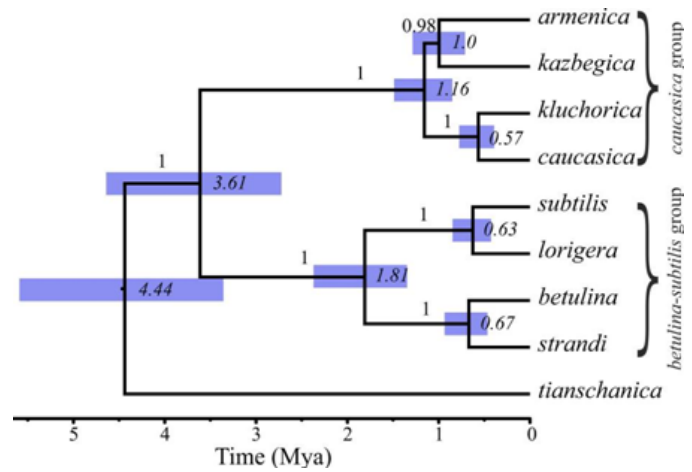


Figure 5 – Species tree / chronogram reconstructed by *BEAST from nuclear and mitochondrial data. Divergence times are shown at nodes (in Mya) with bars representing 95% HPD. Values above branches correspond to posterior probabilities.

Molecular dates

The results of our molecular clock analysis suggest that the radiation among the major birch-mice lineages started in the Early Pliocene or Latest Miocene. The inferred ages of divergence between *caucasica* and *betulina-subtilis* species groups correspond to the Pliocene. Radiation within groups is of the Pleistocene age. These estimates are significantly younger than those obtained in previous molecular studies (Zhang et al., 2013; Pisano et al., 2015). We believe that the latter are in fact an overestimate accounted for by the acceptance of the hypothesis that the earliest fossil *Sicista primus* is the most recent common ancestor of all recent species in the genus. The latter supposition was never justified morphologically. Meanwhile, our estimates do not contradict the fossil data (NOW Database <http://www.helsinki.fi/science/now/database.html>).

Colonization of the Caucasus

The colonization event of the Caucasian region should have happened in the Late Pliocene – Early Pleistocene as follows from our molecular data. Although there are no fossil data of this age on birch mice from the Caucasus, there are reports that at this period *Sicista* sp. are present in Eastern Europe (Nesin and Nadachowski, 2001).

We can assume that birch mice colonized Caucasus region in the Late Pliocene when the ancient Paratethys finally desiccated thus forming a land bridge between Russian Plane and the Caucasus (Popov et al., 2004). At this age Hipparion fauna faces extinction and is replaced by more cold-tolerant species from Eastern Europe (Vereschagin, 1959). In the Early Pleistocene the region saw one of the strongest uplifts, with amplitude around 1500–2500 m (Safronov, 1972). It was followed with further deepening of river valleys. In general, the region was divided into two large river basins: western and eastern with the division in the Central Caucasus. Around 1 Mya active volcanism took place from Elbrus to Kazbek (Safronov, 1972). It may be possible that these events (volcanic activity, rapid uplift and deepening of river valleys) as well as climate change associated with the mid-Pleistocene transition (Mudelsee and Schulz, 1997) promoted new speciation events.

In the Middle and in the Late Pleistocene the Caucasian Mountains were subjected to global glaciations. At least three stages of glaciations are known (Milanovsky, 2008). Glaciation was the strongest in the Central Caucasus but relatively weak in the Eastern and the Lesser Caucasus. Eltübü glaciation (corresponding to Mindel in the Alps) with glaciers descending to about 400–700 m above current river beds may coincide with *kluchorica-caucasica* split but does not necessary explain it. Terek (Riss in the Alps) glaciation was the strongest one with glaciers descending to 180–350 m above current river bed. Notably, Caucasian birch mice are currently unable to survive at elevation below 1400 m asl, animals die quickly if brought down even to 1000 m asl. Nevertheless, birch mice survived glaciations. Potential explan-

ation is that they either survived in small refugia which could exist in some valleys (although none is known so far) or were forced to descend gradually to the foothills and live in the periglacial zone. A rigorous testing of these hypotheses is currently problematic.

Variation within *Sicista tianschanica*

Based on our findings we can conclude that there is another diverse mountain species group of birch mice: *Sicista tianschanica*. The genetic distance between the animal we analyzed from the northern part of the range (the Tarbagatay Mountains) and animals from east Tien Shan (examined by Pisano et al., 2015) is 16% (cytb, k2p). This level is comparable to the distance between “good” species such as *S. betulina* and *S. subtilis*. These mtDNA results are supported by the available data on two nuclear genes. This finding may correlate to karyological variation revealed by Yu. Kovalskaya (Shenbrot et al., 1995). According to her data there exist three probably allopatrically distributed chromosomal forms which are preliminary called as “Terskei”, “Talgar” and “Dzhungar”. The examined here animal from the Tarbagatay range most probably belongs to “Dzhungar” chromosomal race.

It is reasonable to conclude that *Sicista tianschanica* represents an example of a species complex including several cryptic species/subspecies. At the present moment however it appears premature to make any taxonomic decisions, given the lack of information on the holotype of *S. tianschanica*. To conduct a comprehensive revision further study on birch mice of Central Asia is required. ☞

References

- Baker R.J., Bradley R.D., 2006. Speciation in mammals and the genetic species concept. *J. Mammal.* 87(4): 643–662.
- 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.
- Bannikova A.A., Zemlemerova E.D., Colangelo P., Sözen M., Sevindik M., Kidov A.A., Dzuev R.I., Krystufek B., Lebedev V.S., 2015. An underground burst of diversity – a new look at the phylogeny and taxonomy of the genus *Talpa* Linnaeus, 1758 (Mammalia: Talpidae) as revealed by nuclear and mitochondrial genes. *Zool. J. Linn. Soc.* 175(4): 930–948.
- Baskevich M., n.d. *Sicista kazbegica*. Biodiversity Conservation Center. http://www.biodiversity.ru/programs/rodent/species/sicista_kazbegica.html [in Russian]
- Baskevich M.I., Okulova N.M., Potapov S.G., Varshavskii A.A., 2004. Diagnostics, distribution and evolution of unstriped birch mice *Sicista* (Rodentia, Dipodidae) in the Caucasus. *Zool. Zh.* 83(2): 220–233. [in Russian]
- Baskevich M.I., Potapov S.G., Mironova T.A., 2015. Caucasian cryptic species of rodents as models for studying the problem of species and speciation. *Zhurnal Obschej Biologii* 76(4): 319–335. [in Russian]
- Baskevich M.I., Potapov S.G., Mironova T.A., 2016. Caucasian cryptic species of rodents as models in research on the problems of species and speciation. *Biology Bulletin Reviews* 6(3): 245–259.
- Bensasson D., Zhang D.-X., Hartl D.L., Hewitt G.M., 2001. Mitochondrial pseudogenes: evolution’s misplaced witnesses. *Trends Ecol. Evol.* 16: 314–321.
- Cserkészt T., Rusin M., Sramkó G., 2016. An integrative systematic revision of the European southern birch mice (Rodentia: Sminthidae, *Sicista subtilis* group). *Mammal. Rev.* 46(2): 114–130.
- Grigoryeva O., Krivonogov D., Balakirev A., Stakheev V., Andreychev A., Orlov V., 2015. Phylogeography of the forest dormouse *Dryomys nitedula* (Gliridae, Rodentia) in Russian Plain and the Caucasus. *Folia Zool.* 64(4): 361–365.
- Hall T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acid. S.* 41: 95–98.
- Heled J., Drummond A.J., 2010. Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* 27(3): 570–580.
- Ho S.Y., Phillips M.J., Cooper A., Drummond A.J., 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol.* 22(7): 1561–1568.
- Jobb G., 2008. TREEFINDER version of October 2008. Munich, Germany. Distributed by the author at www.treefinder.de.
- Kimura Y., 2013. Intercontinental dispersals of siccistine rodents (Siccistinae, Dipodidae, Rodentia) between Eurasia and North America, In: Wang X., Flynn L.J., Forterius M.

- (Eds.) Fossil mammals of Asia: neogene biostratigraphy and chronology, Columbia University Press. 655–675.
- Lanfeer R., Calcott B., Ho S.Y., Guindon S., 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* 29(6): 1695–1701.
- Lebedev V.S., Bannikova A.A., Pisano J., Michaux J.R., Shenbrot G.I., 2013. Molecular phylogeny and systematics of Dipodoidea: a test of morphology-based hypotheses. *Zool. Scr.* 42(3): 231–249.
- Librado P., Rozas J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11): 1451–1452.
- Milanovsky E.E., 2008. Origin and development of ideas on Pliocene and Quaternary glaciations in northern and eastern Europe, Iceland, Caucasus and Siberia. In: Grapes R.H., Oldroyd D., Grigelis A. (Eds.) History of Geomorphology and Quaternary Geology. Geological Society, London, Special Publications. 301: 87–115.
- Mudelsee M., Schulz M., 1997. The Mid-Pleistocene climate transition: onset of 100 ka cycle lags ice volume build-up by 280 ka. *Earth Planet. Sc. Lett.* 151(1–2): 117–123.
- Nesin V.A., Nadachowski A., 2001. Late Miocene and Pliocene small mammal faunas [Insectivora, Lagomorpha, Rodentia] of Southeastern Europe. *Acta Zoologica Cracoviensia* 44(2): 107–135.
- Pisano J., Condamine F.L., Lebedev V., Bannikova A., Quéré J.P., Shenbrot G.I., Michaux J.R., 2015. Out of Himalaya: the impact of past Asian environmental changes on the evolutionary and biogeographical history of Dipodoidea (Rodentia). *J. Biogeogr.* 42(5): 856–870.
- Popov S.V., Rögl F., Rozanov A.Y., Steininger F.F., Shcherba I.G., Kovac M., 2004. Lithological-Paleogeographic maps of Paratethys. 10 maps Late Eocene to Pliocene. *Cour. For. Sekenbg.* 250: 1–46.
- Rambaut A., Drummond A., 2005. Tracer. A program for analysing results from Bayesian MCMC programs such as BEAST & MrBayes. Version 1.3.
- Ronquist F., Teslenko M., van der Mark P., Ayres D., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P., 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a largemodel space. *Syst. Biol.* 61(3): 539–542.
- Safronov I.N., 1972. Paleogeomorphology of the Northern Caucasus. Nedra, Moscow. [in Russian]
- Sambrook J., Fritsch E.F., Maniatis T., 1989. Molecular cloning: a laboratory manual (No. Ed. 2). Cold spring harbor laboratory press.
- Shenbrot G.I., Sokolov V.E., Goptner V.G., Kovalskaya Yu.M., 1995. Mammals of Russia and adjacent regions. Dipodidae. Nauka, Moscow, Russia. [in Russian]
- Shenbrot G., Bannikova A., Giraudoux P., Quéré J.-P., Raoul F., Lebedev V., 2017. A new recent genus and species of three-toed jerboas (Rodentia: Dipodidae) from China: Alive fossil? *J. Zool. Syst. Evol. Res.* 55(4): 356–368.
- Sokolov V.E., Baskevich M.I., 1988. A new species of birch mice – *Sicista armenica* sp.n. (Rodentia, Dipodidae) from Lesser Caucasus. *Zool. Zh.* 67(2): 300–304. [in Russian]
- Sokolov V.E., Baskevich M.I., 1992. A new chromosomal form of unstriped birch mice of the Caucasus from North Ossetia (Rodentia, Dipodidae, *Sicista*). *Zool. Zh.* 71(8): 94–103. [in Russian]
- Sokolov V.E., Baskevich M.I., Kovalskaya Yu.M., 1981. Revision of birch mice of the Caucasus: sibling species *Sicista caucasica* Vinogradov, 1925 and *S. kluchorica* sp.n. (Rodentia, Dipodidae). *Zool. Zh.* 60(9): 1386–1393. [in Russian]
- Sokolov V.E., Baskevich M.I., Kovalskaya Yu.M., 1986. *Sicista kazbegica* sp.n. (Rodentia, Dipodidae) from the basin of the Terek River upper reaches. *Zool. Zh.* 65(6): 949–952. [in Russian]
- Stephens M., Smith N.J., Donnelly P., 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68(4): 978–989.
- Vereshchagin N.K., 1959. Mammals of the Caucasus. USSR Academic Press, Leningrad. [in Russian]
- Vinogradov B.S., 1925. On the structure of external genitalia in Dipodidae and Zapodidae (Rodentia) as a classificatory character. *P. Zool. Soc. Lond.* 95(2): 582–584.
- Yang Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24(8): 1586–1591.
- Zhang Q., Xia L., Kimura Y., Shenbrot G., Zhang Z., Ge D., Yang Q., 2013. Tracing the origin and diversification of Dipodoidea (Order: Rodentia): Evidence from fossil record and molecular phylogeny. *Evol. Biol.* 40(1): 32–44.

Associate Editor: P. Colangelo

Supplemental information

Additional Supplemental Information may be found in the online version of this article:

Table S1 Specimens used in the study and their accession numbers.

Table S2 Primers used in the study.

Table S3 Descriptive statistics for the genetic markers, summarized for all studied *Sicista* species.

Figure S4 The ML phylogeny of *Sicista* as inferred from BRCA1 and IRBP nuclear genes.