Accepted: 4 February 2019

DOI: 10.1111/jzs.12279

ORIGINAL ARTICLE

WILEY JOURNAL OF ZOOLOGICAL SYSTEMATICS AND EVOLUTIONARY RESEARCH

Phylogeny and evolutionary history of birch mice *Sicista* Griffith, 1827 (Sminthidae, Rodentia): Implications from a multigene study

Vladimir S. Lebedev[1] | Mikhail Yu. Rusin[2] | Elena D. Zemlemerova[3] | Vera A. Matrosova[4] | Anna A. Bannikova[5] | Yulia M. Kovalskaya[3] | Alexey S. Tesakov[6]

[1]Zoological Museum, Lomonosov Moscow State University, Moscow, Russia

[2]Schmalhausen Institute of Zoology, National Academy of Sciences of Ukraine, Kiev, Ukraine

[3]Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow, Russia

[4]Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

[5]Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia

[6]Geological Institute of the Russian Academy of Sciences, Moscow, Russia

Correspondence
Vladimir S. Lebedev, Zoological Museum, Lomonosov Moscow State University, Moscow, Russia.
Email: wslebedev@gmail.com

Funding information
Russian Science Foundation, Grant/Award Number: 14-50-00029; Russian Foundation for Basic Research, Grant/Award Number: 17-04-00065a and 18-04-00400a

Contributing authors: Mikhail Yu. Rusin (mrusin@list.ru), Elena D. Zemlemerova (zemlemerovalena@ya.ru), Vera A. Matrosova (v.matrosova@gmail.com), Anna A. Bannikova (hylomys@mail.ru), Yulia M. Kovalskaya (sicistam@yandex.ru), Alexey S. Tesakov (tesak@ginras.ru)

Abstract

Phylogeny of birch mice is estimated using sequences of ten nuclear genes and one mitochondrial gene. Based on the results of tree reconstructions and molecular dating, five major lineages are recognized: "*tianschanica*," "*concolor*," "*caudata*," "*betulina*," and "*caucasica*." It is established that the three latter lineages constitute a clade and that the long-tailed birch mouse *Sicista caudata* is the sister group of the "*caucasica*" lineage. The "*tianschanica*" lineage is placed as the sister branch to all other species, however, with insufficient support. The *cytochrome b* tree is generally concordant with the nuclear topology. The molecular clock results suggest that the radiation among the main lineages occurred in the Late Miocene–Early Pliocene (6.0–4.7 Mya). The correspondence between molecular dating and the fossil record is discussed. Based on nuclear data, a high level of divergence between cryptic species in the "*tianschanica*" lineage is confirmed. Mitochondrial and nuclear data suggest the existence of a potential cryptic species within *Sicista strandi*.

KEYWORDS

cryptic species, Dipodoidea, molecular clock, Palearctic, multilocus phylogeny, *cytochrome b*, Neogene, Pleistocene, species tree

1 | INTRODUCTION

Sicista Gray, 1827 is the only extant genus of the family Sminthidae, which is the most divergent lineage in the superfamily Dipodoidea (Lebedev et al., 2013; Pisano et al., 2015). Compared to highly specialized obligate bipedal jerboas (Dipodidae) and facultatively bipedal jumping mice (Zapodidae), birch mice are relatively unspecialized. The genus is considered to include 14–17 species (Baskevich, 2016; Holden, Cserkész, & Musser, 2017; Holden & Musser, 2005) distributed in Palearctic where they occur in temperate lowland and mountain forests, steppes, or subalpine shrub and meadows.

The taxonomy of the genus was controversial. Thus, Ellerman and Morrison-Scott (1951) listed only six valid species, while Ognev (1948) recognized nine. Most of the identification keys are based

on the structure of glans penis. Birch mice demonstrate a complex bunodont molar structure, which could potentially provide important clues for group systematics; nevertheless, dental traits are seldom used for discrimination among extant species.

In contrast to phylogenetically close conservative Dipodidae, birch mice demonstrate high level of karyotype variability (Shenbrot, Sokolov, Geptner, & Kovalskaya, 1995; Sokolov, Kovalskaya, & Baskevich, 1982; Sokolov, Kovalskaya, & Baskevich, 1987). Consequently, cytogenetic techniques permitted to describe several new species and support the validity of some of the others (e.g., Sokolov, Baskevich, & Kovalskaya, 1981; Sokolov, Kovalskaya, & Baskevich, 1989).

Based on the combination of morphological and cytogenetic data, several species groups are now recognized within the genus, each including several morphologically similar but chromosomally distinct species.

The *Sicista betulina* group consists of the northern birch mouse *S. betulina* (Pallas, 1779) (temperate forests and taiga from Western Europe to Baikal region) and Strand's birch mouse *Sicista strandi* Formozov, 1931 (southern part of East Europe and North Caucasus). Sometimes (e.g., Baskevich, 2016), the group is believed to also include the gray birch mouse *Sicista pseudonapaea* Strautman, 1949, which is only found in the southwestern part of the Altai region.

The *Sicista subtilis* group is distributed in western and central parts of the Eurasian steppe belt from Hungary to Tuva and west Baikal region. In contrast to other birch mice, members of this group prefer semiarid and arid rather than mesic habitats; thus, the southern birch mouse *S. subtilis* (Pallas, 1773) is the only species which is found in semideserts. More than six karyomorphs were described, and the taxonomic treatment of this variation is controversial with the number of recognized species ranging from three (Cserkész, Rusin, & Sramkó, 2016) to six (Kovalskaya et al., 2011).

The *Sicista caucasica* group is distributed in the subalpine shrub and meadow belt of the Caucasus and includes four allopatric species (Caucasian birch mouse *S. caucasica* Vinogradov, 1925; Kazbeg birch mouse *Sicista kazbegica* Sokolov, Kovalskaya, and Baskevich, 1986; Kluchor birch mouse *Sicista kluchorica* Sokolov, Kovalskaya, and Baskevich, 1980; and Armenian birch mouse *Sicista armenica* Sokolov and Baskevich, 1988), three of which were described based on karyotype.

Several species are not attributed to any species group as their taxonomic relations have been considered uncertain (Baskevich, 2016). The Altai birch mouse *Sicista napaea* Hollister, 1912 is distributed in the northern Altai mountains and east Kazakhstan where it occurs in a wide spectrum of habitats. The long-tailed birch mouse *Sicista caudata* Thomas, 1907 inhabits woodlands on Sakhalin island and the adjacent continental Far East. The Chinese birch mouse *Sicista concolor* (Buchner, 1892) is found in mesic habitats in the mountain ranges at the northeastern and eastern edges of the Tibetan plateau. Poorly known *Sicista leathami* (Thomas, 1893) from Kashmir is often regarded as a subspecies of *Sicista concolor*. The Tian Shan birch mouse *Sicista tianschanica* (Salensky, 1903) is found in the northern and central Tian Shan and the mountain ranges adjacent from the north; it includes three karyomorphs of an unclear taxonomic status (Sokolov & Kovalskaya, 1990). Molecular data (Cserkész et al., 2019; Rusin et al., 2018) also suggest the existence of a cryptic species. The relationships among *S. concolor*, *S. caudata*, and *S. tianschanica* were previously treated controversially, some researchers lumped them under *S. concolor* (e.g., Bobrinsky, Kuznetsov, & Kuzyakin, 1965; Zhang et al., 1997).

Earlier molecular studies (Pisano et al., 2015) confirmed the integrity of the species groups. However, the phylogeny of the genus is not completely resolved due to insufficient sampling of taxa or genes (Cserkész et al., 2019). Molecular data on *S. caudata* are lacking.

In the present study, we aim to elucidate the phylogenetic relationships within *Sicista*, based on the sample of ten nuclear genes and one mitochondrial gene. Next, we estimate the ages of splits among and within species groups, and compare it with available fossil data.

2 | MATERIALS AND METHODS

2.1 | Sampling

We analyzed data on twelve of thirteen currently recognized species of *Sicista*. The original samples include two *S. caudata* stored in the collection of the Zoological Museum of Moscow University. Most of specimens of other species were examined in the previous studies (Pisano et al., 2015; Rusin et al., 2018) but using a smaller sample of genes. In total, 174 sequences (50 animals) were taken from GenBank and 87 sequences (20 animals) were obtained *de novo*. The details for all the material used in the study are given in Appendix A and Supporting Information Table S1.

2.2 | 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, Fritsch, & Maniatis, 1989). We sequenced the complete mitochondrial *cytochrome b* gene (*cytb*, amplicon length 1,242 bp) and fragments of ten nuclear loci: exon 11 of the *breast cancer type 1 susceptibility protein* gene (*BRCA1*, amplicon length 886 bp), exon 1 of the *interphotoreceptor binding protein* gene (*IRBP*, 1,150 bp), exon 10 of the *growth hormone receptor* (*GHR*, 774 bp), *recombination activating protein* genes *1* and *2* (*RAG1*, 1,160 bp; *RAG2*, 954 bp), *cannabinoid receptor type 1* (*CNR1*, 1,056 bp and 1,003 bp), exon 11 of the *breast cancer type 2 susceptibility protein* gene (*BRCA2*, 954 bp), intron 2 of the *thyrotropin* gene (*THY*, 599 bp and 582 bp), intron 13 of the *betaspectrin 1* gene (*SPTBN*, 1,007 bp), and intron 9 of the *protein kinase C* gene (*PRKC*, 606 bp and 721 bp). The information on primers used for amplification and sequencing of *cytb*, *IRBP*, *BRCA1*, *GHR*, *RAG1*, *THY*, *SPTBN*, and *PRKC* was published earlier (Lebedev et al., 2013; Rusin et al., 2018). The nucleotide sequences of the original primers designed for amplification and sequencing

of *RAG2*, *GHR*, *CNR1*, and *BRCA2* are provided in the Supporting Information Table 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 BigDyeTM Terminator Chemistry v. 3.1 (Applied Biosystems, Foster City, CA, USA). The assembling was performed using SeqMan (Lasergene, USA). The sequences obtained in this study were deposited in GenBank (Acc. No.: MK259964–MK259975; MK309823–MK309825, MK309826–MK309828, MK259977–MK259979, MK323046–MK298389), details see in the Supporting Information Table S1).

2.3 | Tree reconstruction and molecular dating

All sequences were aligned by eye using BioEdit v. 7.0.9.0 (Hall, 1999). Numbers of variable characters and consistency index values (based on maximum parsimony trees) were calculated in PAUP* 4.0b10 (Swofford, 2003). Base frequency homogeneity was tested in 3rd codon positions using disparity index test (1,000 replicates) as implemented in MEGA 6 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). To correct for multiple comparisons, we applied the false discovery rate procedure (Benjamini & Hochberg, 1995) with the false discovery rate (FDR) kept at 0.05. Genetic distances (Kimura two-parameter (K2P) and uncorrected p-distances) were calculated in MEGA 6. Alignments are provided as Supporting Information.

The concatenated nuclear alignment comprised sequences of three introns and seven exons (8,773 bp in total) and included 32 operational taxonomic units (OTUs) of *Sicista*. The outgroup included representatives of five other taxa of Dipodoidea: two lineages of Zapodidae (*Eozapus* and *Zapus* or *Napaeozapus*) and three subfamilies of Dipodidae–Dipodinae (*Dipus* or *Jaculus*), Allactaginae (*Paralactaga*), and Cardiocraniinae (*Cardiocranius*). The sequences corresponded to unphased genotypes with heterozygous position coded using the IUB ambiguity codes.

Phylogenetic reconstructions were conducted with the following datasets: (a) ten nuclear genes combined; (b) the *cytb* alignment; (c) the combined nuclear and mitochondrial concatenation. Separate analyses of the nuclear and mitochondrial alignments were performed in order to identify potential inconsistencies between the two sets of data, which could arise due to introgression or compositional heterogeneity and saturation in the *cytb* sequences.

Phylogenetic trees were inferred under maximum likelihood (ML) and Bayesian criteria. Maximum likelihood reconstructions were conducted in IQ-TREE version 1.6 (Nguyen, Schmidt, von Haeseler, & Minh, 2015). The ModelFinder routine (Kalyaanamoorthy, Minh, Wong, von Haeseler, & Jermiin, 2017) was used to determine the optimum partitioning scheme and the best-fit substitution models for each subset under BIC. Clade stability was tested using Ultrafast Bootstrap (Minh, Nguyen, & von Haeseler, 2013) with 10,000 replicates.

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 using ModelFinder. For most parameters, default priors were used. Compound Dirichlet priors for branch lengths combined with gamma prior on the tree length were invoked. All parameters except branch lengths were unlinked across partitions. The analysis included two independent runs of four chains with the default heating scheme. The chain length was set at 20 million generations with the sampling of every 10,000 generation. Tracer 1.6 software (Rambaut & Drummond, 2003) 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.

Additional Bayesian analyses were performed on each of the separate gene alignments employing the same set of priors as with concatenated data. The chain length was set at 11 million steps, and the first million generations were discarded as burn-in.

A more complete mitochondrial phylogeny was reconstructed based on the extended *cytb* alignment including 42 sequences of *Sicista*. The ML and Bayesian analyses were conducted using the same algorithms and procedures as with the nuclear data. In all reconstruction, the *cytb* dataset was subdivided into three subsets corresponding to codon positions. Saturation was evaluated by plotting the *cytb* distances (K2P) against corresponding nuclear distances.

As a complement to concatenation-based reconstructions, we used two multilocus approaches designed to estimate a species tree from a sample of potentially incongruent genes. First, the species tree was reconstructed employing a Bayesian coalescent framework as implemented in *BEAST (Heled & Drummond, 2010). Seventeen lineages and sublineages of *Sicista* and five outgroups were used as OTUs. If available, additional sequences were included in the individual gene alignments to improve the accuracy of the population parameter estimation. Prior to analysis, the genotype data on each of the 10 nuclear genes were phased using Phase software (Stephens, Smith, & Donnelly, 2001) via DNAsp ver. 5 (Librado & Rozas, 2009). Preliminary runs with default priors showed that stationarity is not reached even after 500 million generations. To facilitate convergence, no calibration information was used, clock rate for *BRCA1* was fixed at unity, and prior densities for clock rates of other genes were modeled using Gamma distributions with scale and shape parameters estimated from posterior distributions inferred from the concatenated data in BEAST ver. 1.8.4 (Drummond, Suchard, Xie, & Rambaut, 2012; see below). 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. Four runs of one billion generations each were conducted saving every 500,000th generation. Parameter convergence was assessed in Tracer 1.6 software (Rambaut & Drummond, 2003).

As a second approach, the supertree was inferred with matrix representation with parsimony (MRP; Baum, 1992; Ragan, 1992) using posterior samples of individual gene trees reconstructed by MrBayes as input. In contrast to *BEAST, this method does not assume that incomplete lineage sorting due to ancestral polymorphism

is the only cause of discordance among gene trees and species trees. The analysis was performed in PAUP* 4.0b10 (Swofford, 2003), and clade support was estimated with the use of the procedure described in Bannikova, Lebedev, Abramov, and Rozhnov (2014).

Molecular dating was performed based on nuclear concatenation. This approach was chosen because of the difficulties encountered in species tree reconstruction in *BEAST that are mentioned above. The *cytb* data were not included in this analysis because the estimates of divergence based on this gene were substantially biased downward for more ancient lineages, which is likely an effect of saturation. The hypothesis of strict clock was tested for each gene using hierarchical likelihood ratio tests using PAML ver 4.7 (Yang, 2007) based on the ML topology. To verify that there is no rate variation between Sminthidae and Dipodidae + Zapodidae, the analyses were performed with the inclusion of additional outgroups from Spalacidae and Cricetidae. The ultrametric calibrated tree (chronogram) was reconstructed in BEAST ver. 1.8.4 (Drummond et al., 2012). Partitioning and substitution models corresponded to those in the ML analysis. We used the Bayesian skyline model (10 groups) as a flexible tree prior. Separate strict or relaxed lognormal clock model was selected for each gene depending on the result of the molecular clock test. Uniform distribution between 0 and 0.1 was used as prior for clock rates. Default priors were used for all other parameters. The tree was calibrated using two secondary calibrations that follow from the results of the Bayesian dating by Shenbrot et al. (2017): the time of split between Allactaginae and Dipodinae (~17.3 Mya, 95%HPD: 18.7–16.2) and the time of separation of Cardiocraniinae from other dipodids (~20.4 Mya, 95%HPD: 23.1–18.1). Prior calibration densities were defined using lognormal distribution. The chain length was 100 million generations, and parameters were logged to file every 50,000th step. The data on fossil *Sicista* were not used for tree calibration because of the ambiguities in assignment of fossil taxa to extant lineages as it is argued by Rusin et al. (2018). Thus, the cladistic analysis of dental characters in recent, Pleistocene and Neogene *Sicista* (Kimura, 2013) produced a poorly resolved tree with no well-supported associations between extant and fossil taxa.

2.4 | Ancestral area estimation

To reconstruct the history of the genus *Sicista*, an ancestral area estimation was performed with the use of the maximum likelihood and maximum parsimony algorithms implemented in BioGeoBEARS ver. 1.1 (Matzke, 2013) and DIVA module of RASP ver. 4.0 (Yu, Harris, Blair, & He, 2015), respectively. We defined eight biogeographic areas based on the distribution of birch mice species: Europe, Caucasus, Siberia, Kazakh steppe, Tian Shan and Altai mountain regions, Qinghai–Tibet plateau and adjacent mountains, Mongolia and Northern China, and Far East. Maximum range size was fixed to four areas. Constraints on dispersal between areas were imposed either using the adjacency matrix or by explicit specification of allowed ancestral ranges, which included all non-disjunct ranges consisting of four or less areas. The analyses were performed using the time-calibrated species tree inferred in *BEAST. In the ML reconstructions, four models of range evolution (DEC, DEC+J, DIVALIKE, and DIVALIKE+J) were employed. The models were compared for statistical fit using Akaike information criterion (AIC). The dataset included fossil *Sicista prima* (Early Miocene, N China), which is placed as the sister taxon to all recent *Sicista* as follows from Kimura (2013). Other fossil taxa cannot be included due to the uncertainty of their phylogenetic relationships. More details on the analyses in BioGeoBEARS and RASP are given in Supporting Information ST1.

2.5 | Species validation

To test for significant genetic differentiation between the two cryptic sublineages in *S. strandi*, we employed a multilocus species validation method implemented in the BPP 3.4 software (Yang & Rannala, 2010). The analysis was performed based on the nuclear dataset using the guide species tree containing a single ancestral node. The relative rates of the loci were fixed at values inferred by BEAST. The inverse-gamma priors, IG (3, 0.004) and IG (3, 0.006), were specified for θ and τ parameters, respectively. Several runs using different rjMCMC algorithms were performed as described in Yang and Rannala (2010). The MCMC chain length was 200 thousand generations, and the burn-in period was 100 thousand generations.

3 | RESULTS

3.1 | Phylogenetic trees

Following the results of ModelFinder, the nuclear concatenation was partitioned by gene and codon position (24 subsets in total; for models, see Supporting Information Table S3). The trees inferred from the nuclear alignment with ML and Bayesian methods showed the same topology (Figure 1). The *cytb* dataset was less informative than the nuclear concatenation (522 vs 2110 variable characters; Supplementary Information Table S4) and showed more homoplasy (consistency index values are 0.36 and 0.88, respectively). Substantial saturation at the deeper nodes was revealed (Supporting Information Figure S1). The mitochondrial tree (Figure 2), however, generally agreed with the nuclear phylogeny as it contained no well-supported clades in conflict with the latter. Correspondingly, the tree reconstructed from the combined nuclear + mitochondrial concatenation was almost identical to the nuclear tree and had higher support for most of the nodes (Figure 3).

The relationships among Sminthidae and other families and subfamilies of Dipodoidea followed the pattern recovered in the previous studies (Lebedev et al., 2013; Pisano et al., 2015). All species of *Sicista* clustered into five well-supported major lineages: "*betulina*," "*caucasica*," "*caudata*," "*tianschanica*," and "*concolor*."

The "*betulina*" lineage demonstrated complex internal structure including three subclades corresponding to *S. subtilis* species group (*S. subtilis*, *S. nordmanni*, and *S. trizona*), *S. betulina* species group

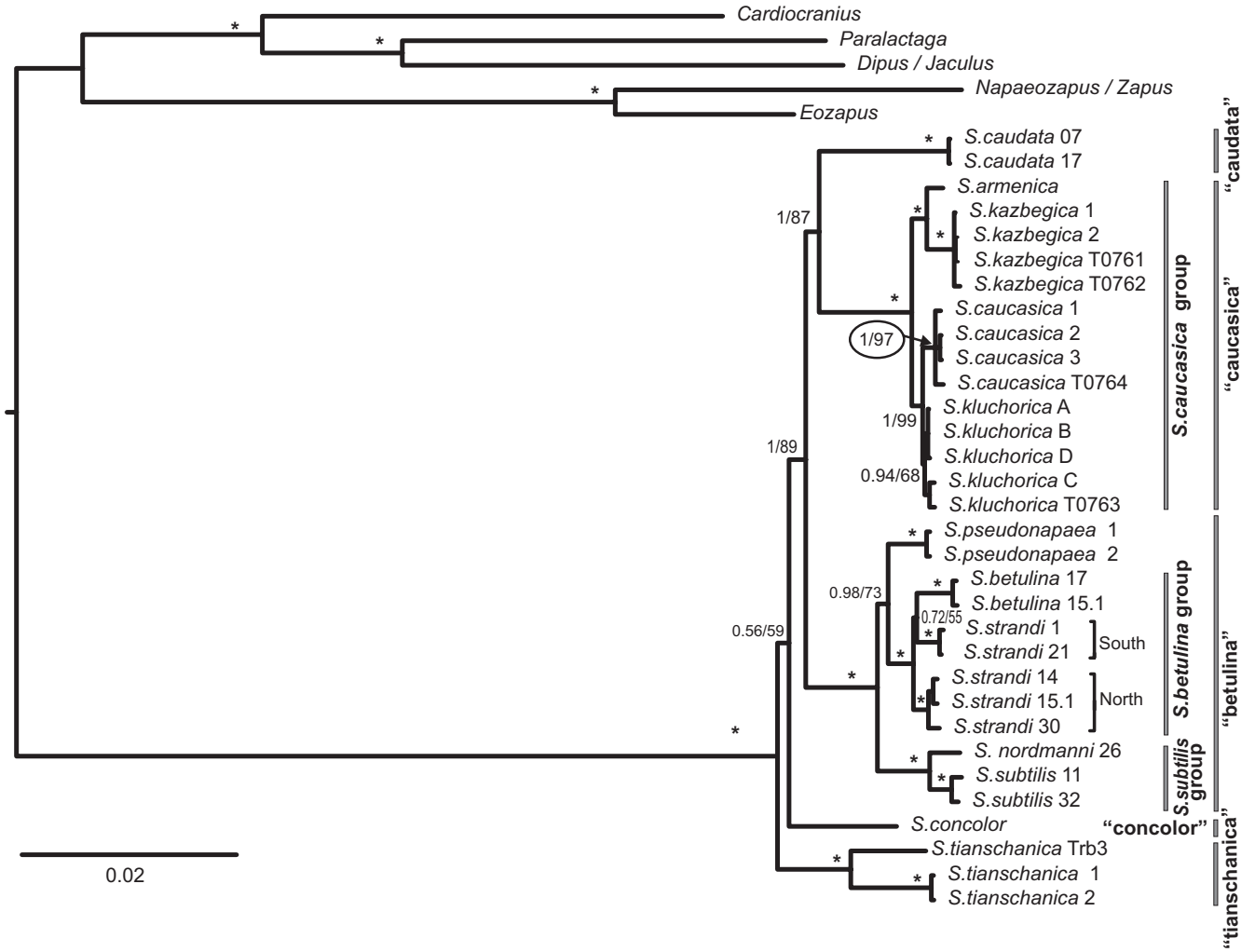

FIGURE 1 The phylogenetic relationships in the genus *Sicista* as reconstructed in MrBayes from the concatenation of ten nuclear genes. Numbers above or below branches correspond to Bayesian posterior probabilities and ML bootstrap values generated using fast bootstrap algorithm in IQ-TREE. The tree is rooted using Dipodidae and Zapodidae as the outgroup. Asterisks denote posterior probabilities of 1.0 and bootstrap values of 100%. Bootstrap support is only shown for the values exceeding 50%. Support values for intraspecific relationships are mostly omitted. Genetic lineages (in quotes) and species groups are shown

(*S. betulina* and *S. strandi*), and *S. pseudonapaea*, which was placed as the sister branch of the *S. betulina* group. However, the support for this position ranged from low to high.

The pattern inferred for the "*caucasica*" lineage confirmed sister group relationships between *S. caucasica* + *S. kluchorica* and *S. armenica* + *S. kazbegica* clades, as it was previously demonstrated (Rusin et al., 2018).

In the nuclear trees, the basal split separated the "*tianschanica*" lineage from remaining species; however, the support for this arrangement was generally low, indicating a potential trichotomy involving also *S. concolor*. The "*betulina*," "*caucasica*," and "*caudata*" lineages constituted a clade that was supported by the nuclear data (PP = 1.0, BS = 90%) as well as by the nuclear + mitochondrial concatenation (PP = 1.0, BS = 98%). Within the latter association, the "*caucasica*" and "*caudata*" lineages were recovered as sister groups with high to moderate support (PP = 1.0, BS = 89%–91%).

The Bayesian analyses of the individual gene alignments produced trees with various degree of resolution (Supporting Information Figure S2). The results suggest that some of the gene trees were not fully concordant with the topology inferred from the concatenation. Thus, the *IRBP* and *RAG2* trees contained a well-supported association of the "*betulina*" and "*caucasica*" lineages, while in the *RAG1* tree, *S. caudata* was placed as the sister group to the remaining species.

At the same time, the phylogenetic trees inferred by the two multilocus methods were fully concordant with the results of the concatenated analysis (Figure 4a, b). We have to remark that the species tree reconstruction in *BEAST suffered from slow mixing, which can be attributed to the impact of missing data. Nevertheless, the effective sample size exceeded 200 for the majority of parameters including the likelihood and prior. All four runs generated the same MCC topology with similar values of posterior probabilities for the supported clades.

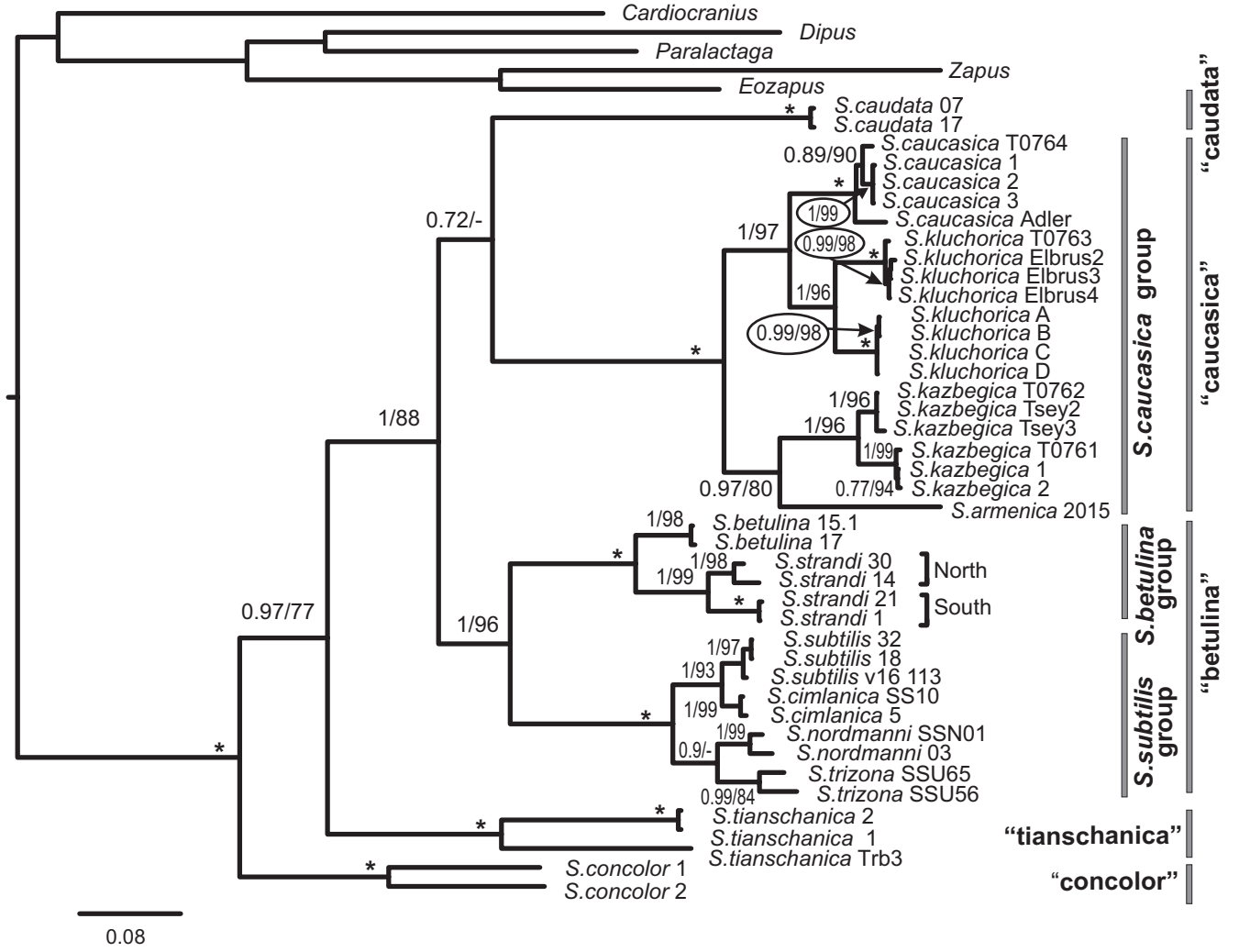

FIGURE 2 The phylogenetic relationships in *Sicista* as reconstructed in MrBayes based on the mitochondrial *cytb* gene alignment. Support values are denoted as in Figure 1

It should be mentioned that, as followed from the results of the pattern homogeneity tests, the *cytb* gene demonstrated substantial variation in base composition, the *S. betulina* and *S. subtilis* species groups being significantly more AT-rich than the remaining species (with and without FDR correction). Therefore, the results of the *cytb*-based reconstructions should be taken with certain caution. In contrast to that, neither of nuclear gene demonstrated significant (after FDR correction) departure from homogeneity within *Sicista*. The detailed results of the base homogeneity tests are available from the authors.

3.2 | Cryptic taxa

The *cytb* distances between currently recognized species of *Sicista* range between 7.3% and 26%, while distances calculated from nuclear concatenation vary between 0.3% and 3.4% (Supporting Information Tables S5 and S6). Both nuclear and mitochondrial data indicated that *S. strandi* includes two well divergent sublineages (*cytb* K2P distance 6%). The first one occurs in the Lower Don and Volga valleys, while the second is only found in Belgorod–Kursk area. The level of nuclear divergence between them (0.5%, K2P) was similar to that between *S. betulina* and *S. strandi* (0.6%). The results of the BPP analysis supported recognition of the two sublineages of *S. strandi* as separate entities ($P = 1.0$) regardless of the algorithm.

In addition, our nuclear data supported the high level of divergence between the two species belonging to the "*tianschanica*" clade (1.4%, K2P). The mitochondrial results indicated a likely existence of cryptic sublineages in *S. concolor* (15% of *cytb* K2P distance between the two sequences, both from Sichuan, China).

3.3 | Molecular dating

The results of molecular dating are presented in the Supporting Information Table S7 and Figure 5. For all genes except *CNR1* and *RAG1*, likelihood ratio tests did not reject

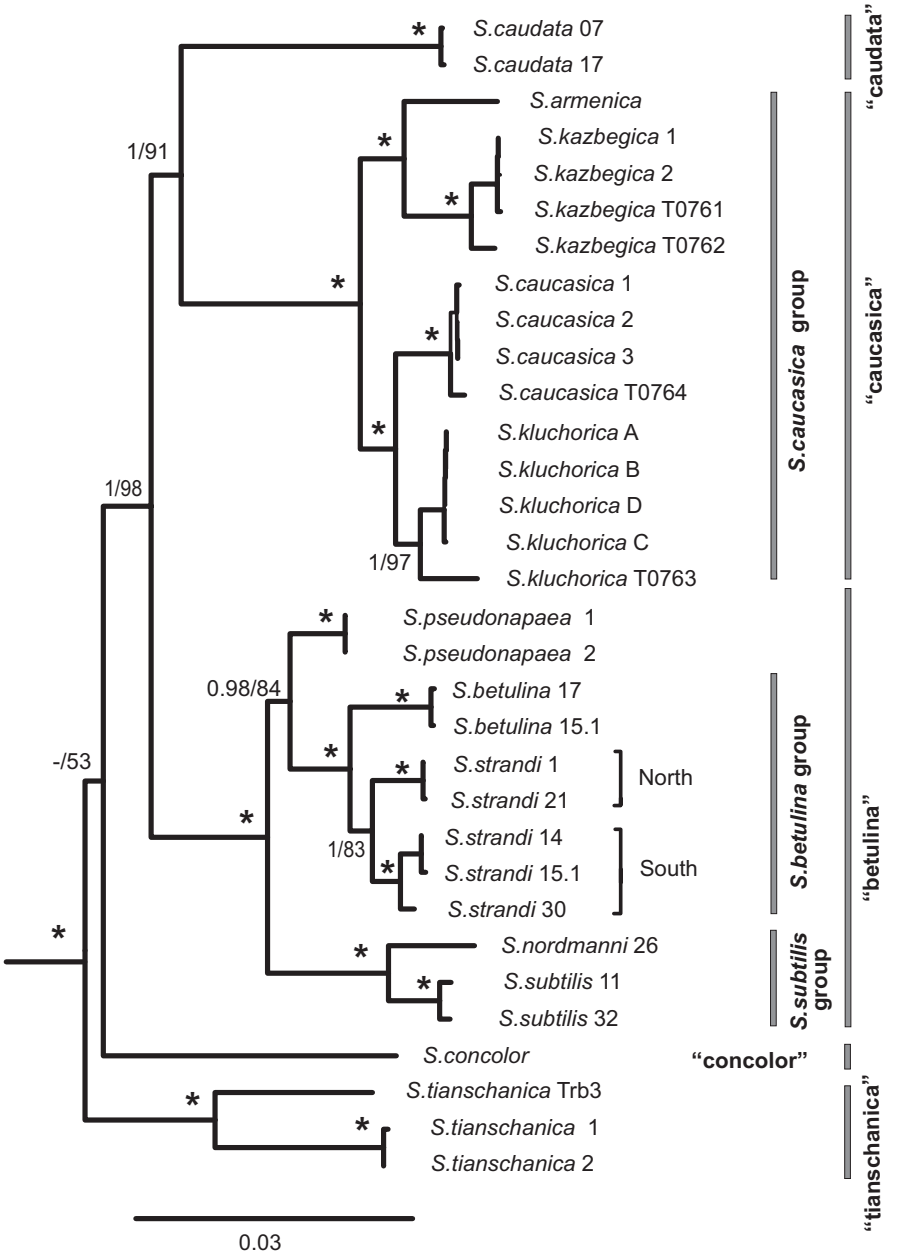

FIGURE 3 The phylogenetic relationships in the genus *Sicista* as reconstructed from the concatenation of the nuclear and mitochondrial datasets. The topology and branch lengths correspond to the ML tree inferred in IQ-TREE. Support values are denoted as in Figure 1. The outgroup (Dipodidae and Zapodidae) is not shown

the strict clock model. In case of *RAG1*, the assumption of rate constancy held for all Dipodoidea if a local clock model implying a separate rate for the *Zapus–Napaeozapus* branch was employed. In the *CNR1* data, no particular branch responsible for clock violation was identified. Therefore, we performed the molecular clock analysis and species tree inference with the exclusion of *RAG1* sequences of the *Zapus* lineage and assuming the relaxed clock model for *CNR1*. The topology of the chronogram (Figure 5) was consistent with that of the other phylogenetic trees. The ages of the splits among the five main lineages fell within the interval between 6.0 and 4.7 Mya (Late Miocene–Early Pliocene). The divergence between the two branches in the *S. tianschanica* species group was dated back to Late Pliocene (~3.0 Mya). The time of the most recent common ancestor of the "*betulina*" lineage, which corresponds to the separation of the *S. subtilis* species group, was placed at the Pliocene–Pleistocene boundary (~2.6 Mya). The times of the separation of allopatric taxa within the "*caucasica*" lineage, *S. subtilis* and *S. betulina* species groups, were close to ~1 Mya.

3.4 | Ancestral area estimation

Ancestral area reconstructions yielded ambiguous results (see Supporting Information ST1). In most cases, the reconstructed range of the common ancestor of the crown *Sicista* included Tian Shan–Altai region combined with Qinghai–Tibet region or Siberia. The maximum likelihood estimates of the ancestral

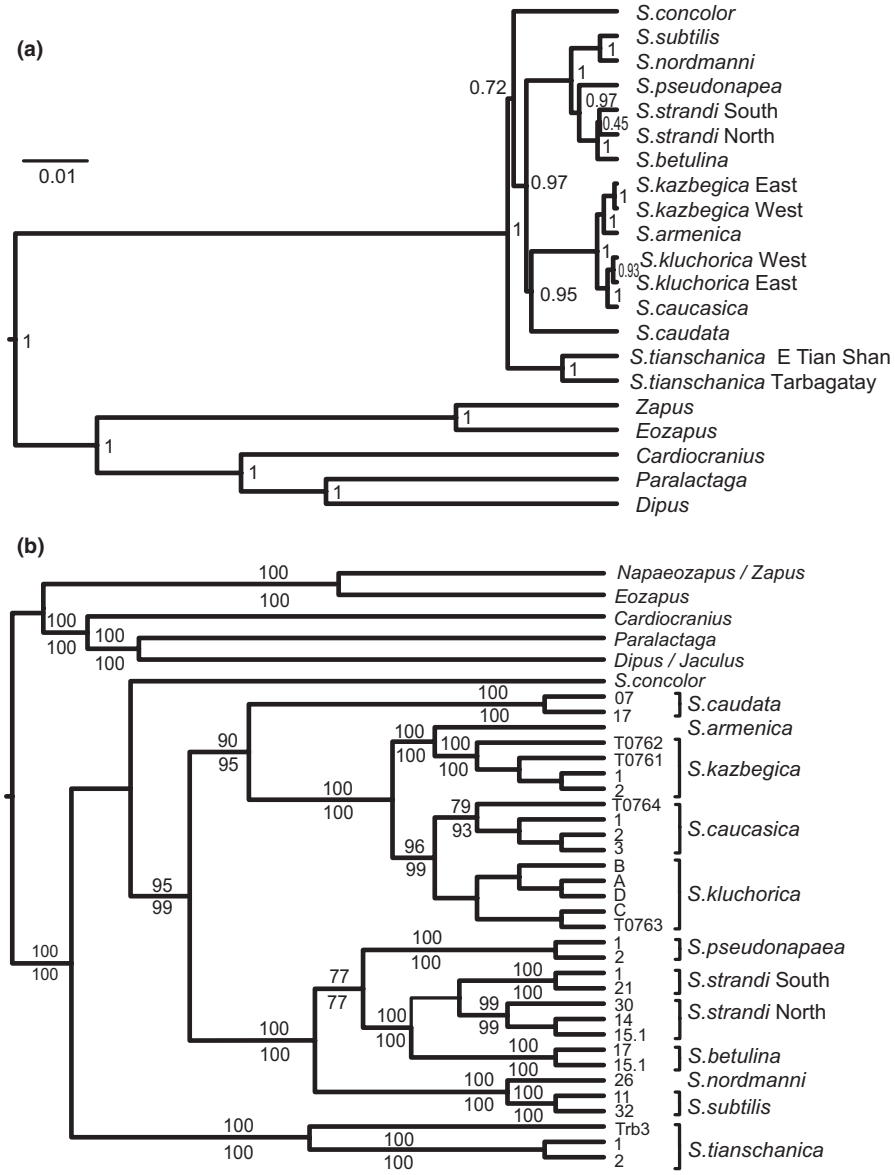

FIGURE 4 The results of phylogenetic reconstructions using multilocus methods. (a) The species tree generated in *BEAST, values above branches correspond to posterior probabilities. (b) The consensus tree reconstructed using MRP method. Numbers below branches denote support values calculated from posterior distributions of eleven individual gene tree topologies produced by MrBayes; the number above branches correspond to values calculated with the *cytb* data excluded

ranges for the main clades were found to be sensitive to model assumptions and particularly to dispersal constraints. This failure to produce consistent results may be due to the fact that the data on a relatively compact group can be insufficient for accurate estimation of model parameters within the likelihood framework (Pirie, Humphreys, Antonelli, Galley, & Linder, 2012). At the same time, parsimony reconstructions are known to be prone to bias due to oversimplified assumptions of the method (Kodandaramaiah, 2010). Moreover, all available biogeographic models may be too unrealistic for reconstruction of complex evolutionary histories such as that in birch mice. To illustrate the complex nature of range evolution in *Sicista*, one should note that the contemporary range of the genus does not include its primary center of origin (see Discussion section). Therefore, our discussion of the evolutionary history of birch mice is based on the fossil evidence and the molecular clock analysis.

4 | DISCUSSION

4.1 | Molecular phylogeny and molecular dating of the genus *Sicista*

Resolving the branching order in *Sicista* appears a challenging task due to a nearly bush-like radiation among the main lineages and the lack of a close outgroup. Thus, the stem of the birch mouse clade is 4.5 times longer than the root-to-tip distance of crown *Sicista*, the latter equals ca. 6 Mya as follows from the molecular clock results. At the same time, the divergence events among the five main lineages occurred over the time span of just 1.5 Mya (from 6.0 to 4.7 Mya). In some cases, individual genes supported conflicting topologies, emphasizing the necessity for a multigene analysis. The sampling of ten nuclear genes used in this study proved to provide adequate resolution and support for most of the nodes.

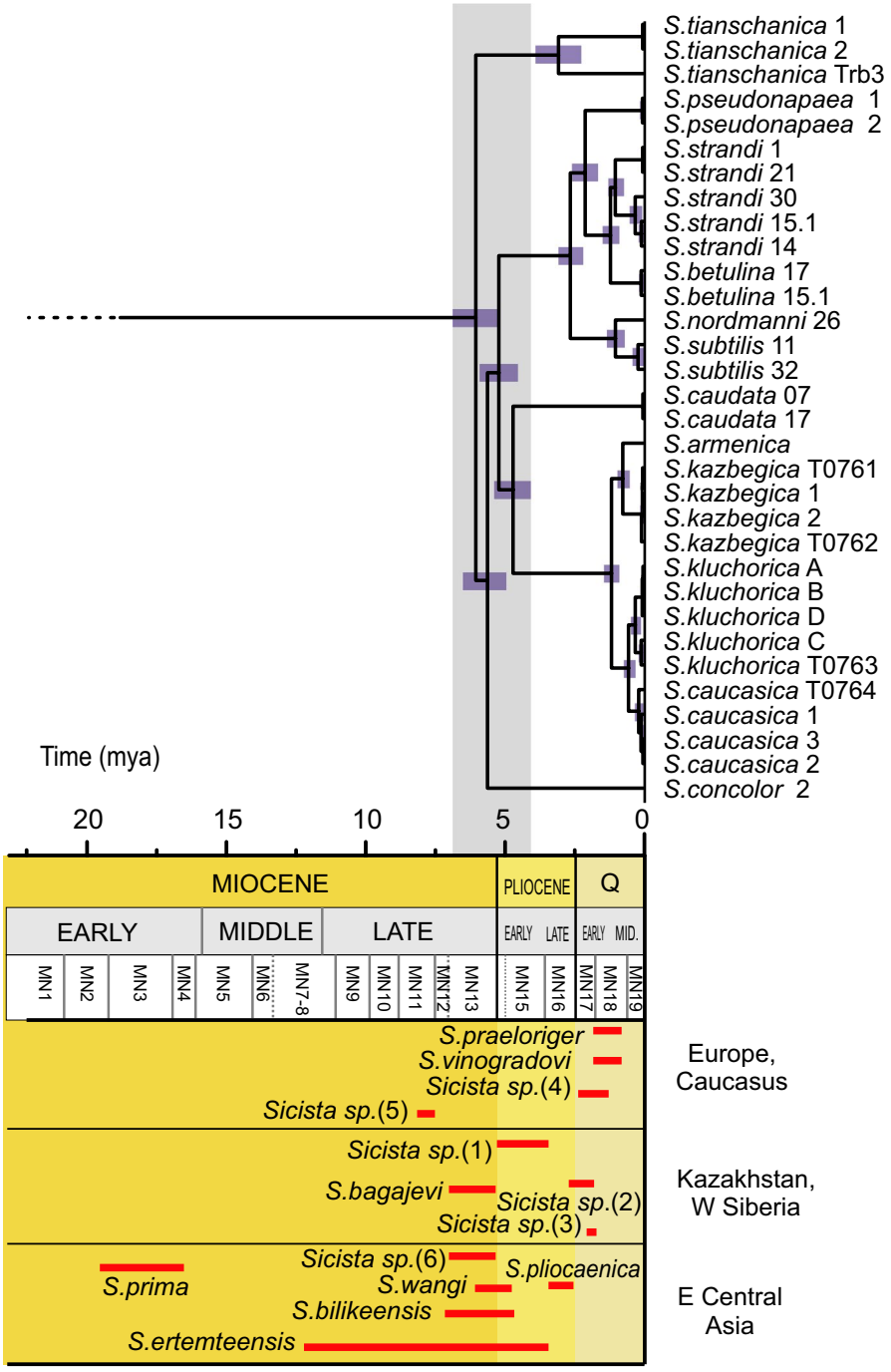

FIGURE 5 The chronogram illustrating major divergence events among *Sicista* as inferred from nuclear concatenation in BEAST. The node bars represent the 95% HPD intervals of the node ages. The vertical gray band indicates the time of rapid cladogenesis in *Sicista* as follows from the molecular results. Below is the chart representing the timescale of fossil record of *Sicista*. The references to *Sicista spp*. are as follows: 1—Rytovka Formation, Biteke Formation (Zykin, 2012); 2—Podpusk-Lebyazhie (Zazhigin, 2009); 3—Kopaly (Tjutkova & Kaipova, 1996), Isakovka 4 (Tesakov et al., 2016); 4—(Tesakov, 2004; Topachevsky et al., 1987); 5—Gaverdovsky Formation (Tesakov et al., 2017); 6—Yushe (Qiu, 2017)

While the inferred pattern of phylogenetic relationships considers more genes and taxa and is thus more informative, its results are generally congruent with those obtained in the previous studies (Cserkész et al., 2019; Pisano et al., 2015; Rusin et al., 2018). The data firmly support the existence of five main lineages, one of which ("*caucasica*") corresponds to a recognized species group.

No molecular data on the long-tailed birch mouse *S. caudata* were previously available. Our results indicate that this species is a single representative of a separate ancient lineage. It is not related to either *S. concolor* or *S. tianschanica* as suggested earlier (Bobrinsky et al., 1965; Corbet, 1978). In contrast, the long-tailed birch mouse is found to be a distant sister group of the *S. caucasica* species group, but the length of the common stem is rather short. Paradoxically, the members of these two lineages are distributed at the opposite sides of the genus range—southwest for "*caucasica*" versus northeast for "*caudata*." They also inhabit contrasting habitat types—high-altitude subalpine shrubs and meadows versus lowland conifer forests, respectively. The molecular evidence is inconsistent with the supposition that *S. caudata* is the most ancient species of birch mice, as it was argued by Baskevich (2016) based on its supposedly plesiomorphic karyotype (2*n* = 50, NFa = 48; Sokolov et al., 1982). Instead, the available nuclear data recapitulate earlier results (Pisano et al., 2015), indicating the basal position of the "*tianschanica*" lineage, followed by the "*concolor*" lineage. However, while the clade including

"*caudata*," "*caucasica*," and "*betulina*" lineages is well supported, the support for the placement of *S. tianschanica* as the sister group to the rest of *Sicista* is generally low. This lack of resolution may be explained by a smaller number of genes examined in *S. concolor*. Apparently, the hypothesis of basal position for *S. concolor* cannot be rejected as of today. The latter species is still poorly known; the mitochondrial data available indicate significant differentiation among the Chinese *S. concolor* with the distance between the two haplotypes (15%) being consistent with species-level differentiation (>11%) according to Bradley and Baker (2001). The status of these genetic sublineages as well as the relationships of the nominotypical form (Gansu, China) with *S. c. weigoldi* (Sichuan, China) and *S. c. leathami* (Kashmir, India) remains to be clarified. The latter taxon is isolated from the others by a large distance as well as substantial geographic barriers and may, in fact, be a distinct species.

Nuclear data confirm that the *S. tianschanica* group contains at least two highly divergent sublineages as it was previously shown by Cserkész et al. (2019) and Rusin et al. (2018). Taking into account the old age of the split between these taxa (~3 Mya), one can conclude that they may deserve species rank; however, their distribution is unclear as of yet. Earlier (Shenbrot et al., 1995; Sokolov & Kovalskaya, 1990), *S. tianschanica* was found to include three karyomorphs, which were preliminary named as "Terskei" (2*n* = 32, NFa = 56) occurring in central Tian Shan, Kungey Alatau, Terskey Alatau, and Ketmen, "Djungar" (2*n* = 34, NFa = 54) distributed in Dzungar Alatau, Tarbagatai, Saur, and "Talgar" (2*n* = 32, NFa = 56, with the chromosome morphology being different from "Terskei"), which was found in two localities in Trans-Ili Alatau and Dzungar Alatau. Cranial differences among the karyomorphs have been discussed (Shenbrot et al., 1995). One may expect that the specimen from Tarbagatay belongs to "Djungar" morph. However, the affinities of the specimens from eastern Tian Shan examined by Pisano et al. (2015) as well as of those collected in Dzungar Alatau and Trans-Ili Alatau by Cserkész et al. (2019), are uncertain. Taking into account the high level of divergence between the lineages, the potential sympatry of cryptic species cannot be ruled out. Therefore, although the collecting locality in eastern Tian Shan (Narati) is relatively close to the terra typica (Hapzagai-gol, upper Ili valley), it remains unclear which phylogenetic lineage corresponds to typical *S. tianschanica*. Based mostly on genetic data for several specimens from Dzungar Alatau, Cserkész et al. (2019) described a new species *Sicista zhetysuica* and assumed that they belong to "Djungar" karyomorph. However, they failed to present supporting chromosomal evidence and did not compare their sample with the type series of *S. tianschanica*. Consequently, we believe that the status of *S. zhetysuica* requires confirmation, while the relationships among genetic sublineages, morphological groups, and karyomorphs within the *S. tianschanica* species group should be further examined based on a larger sample.

The pattern of relationships within the "*caucasica*" group follows the one reconstructed by Rusin et al. (2018), which highlights the position of *S. armenica* as the sister branch of *S. kazbegica*.

In contrast to other major clades, the "*betulina*" lineage includes more than one species group (*S. betulina* group, *S. subtilis* group, and, probably, *S. napaea*). *S. pseudonapaea*, which is often regarded as a member of the *S. betulina* species group (e.g., Baskevich, 2016), is placed here as a relatively distant sister group of the latter. The analysis of one nuclear and one mitochondrial gene by Cserkész et al. (2019) suggests that *S. pseudonapaea* is closely related to *S. napaea*, which is usually treated as a species of uncertain taxonomic position due to its specific penile morphology. Having considered these facts, we believe that *S. pseudonapaea* should not be included in the "*betulina*" species group, rather it should be attributed to a separate species group together with the Altai birch mouse. A high level of variation in glans penis morphology observed among species of the "*betulina*" clade (for details see Shenbrot et al., 1995) indicates an elevated rate of penile evolution in this lineage.

Our molecular genetic analysis definitely supports species-level divergence between *S. betulina* and *S. strandi*. The status of the latter species was questioned by Cserkész et al. (2019) due to the small genetic distance from *S. betulina*. However, their result is likely an artifact stemming from the inclusion of *cytb* pseudogene sequences as discussed in Rusin et al. (2018). Moreover, our data reveal two cryptic sublineages within *S. strandi*, which are provisionally designated here as Southern (Lower Don Valley, the right bank of the Volga, southeastern Ukraine) and Northern (Belgorod region—the central part of the East European "chernozem belt"). The level of *cytb* divergence between these phylogroups falls within the range characteristic for both inter- and intraspecies variation (2%–11%) following Bradley and Baker (2001). According to the nuclear data, the two sublineages of *S. strandi* and *S. betulina* stand in an unresolved trichotomy, which dates back to the end of the Early Pleistocene (ca. 1 Mya). In most of nuclear loci, the northern and southern sublineages share no common alleles. Taken together, these facts suggest that the two sublineages may represent distinct species. The analysis of mtDNA variation (Baskevich et al., 2015) showed that the "Southern" mitochondrial haplotypes are found also in the North Caucasus (Ossetia) thus suggesting that the Southern phylogroup may correspond to the nominotypical *S. strandi*, which was described from the Great Caucasus (Karachay region). In this case, the Northern sublineage should be regarded as an undescribed taxon. However, this supposition should be validated by an examination of the material from type locality and the holotype. Morphological diagnoses for the two sublineages require an additional study. Previous research that focused on morphological and chromosomal variation within *S. strandi* demonstrated certain differentiation between northern and southern populations (Baskevich, Okulova, Oparin, & Vlasov, 2005). However, the geographical sample used therein was insufficient for reliable conclusions. The two sublineages may differ in the structure of glans penis as it is illustrated in Shenbrot et al. (1995). In the birch mice from the North Caucasus, the size of spines covering the lateral surface of glans penis is much larger than that in the animals from the northwest of the range; yet, it remains to be established whether this feature can be viewed as a diagnostic trait.

Although chromosome data contributed substantially to our understanding of birch mice systematics, the details of karyotype

evolution in the genus are still mostly unknown. The chromosome complement is known for all lineages and species except *S. concolor* (Baskevich, 2016; Shenbrot et al., 1995). However, as it is illustrated by the analysis of high-quality G-banding data in the *S. subtilis* group (Kovalskaya et al., 2011), the spectrum of chromosome rearrangement in *Sicista* includes not only Robertsonian fusions but also tandem fusions and pericentric inversions. This produces substantial variation in fundamental number (NFa) and complicates the identification of homologous blocks of synteny. Taking into account the deficit of comparative chromosome painting (FISH) data on *Sicista*, its ancestral karyotype can hardly be inferred accurately. It might be tentatively hypothesized that the ancestral NFa for the "*betulina*," "*caucasica*," and "*caudata*" lineages is 48 or 50 as these numbers are observed in *S. kazbegica*, *S. armenica*, *S. strandi*, *S. napaea*, *S. pseudonapaea*, *S. caudata*, while, in the *S. tianschanica* group, NFa is larger (54—56). The rate of chromosomal evolution is the highest within the "*betulina*" lineage as it includes species with both the highest and lowest NFa: 60 in *S. betulina* and 26 in *S. (subtilis) sp2* sensu Kovalskaya et al. (2011).

4.2 | Fossil history of *Sicista*

To test the validity of our reconstructions, one should assess the consistency between the molecular divergence times and the available fossil data, which is briefly reviewed in this section. The Late Oligocene–Early Miocene radiation of primitive Dipodoidea produced multiple genera of the group in Eurasia and North America. However, it is not until Early Miocene that the dental morphology of *Sicista* appeared in the fossil record. *Sicista* is among very few extant rodent genera that have such a long geological history. The earliest obvious *Sicista* specimens were found in Early Miocene faunas of North China dated as latest Xieijan to Shanwangian Land Mammal Ages, rough equivalent of MN2-3 units of the European biochronology, ca. 20–18 Mya (Deng, Hao, Guo, & Zhu, 2019; Qiu & Li, 2016). The earliest birch mouse, *S. prima* Kimura, 2011 shows the smallest dimensions and relatively simple structure of upper molars (Kimura, 2011, 2013; Qiu & Li, 2016). In the cladistic analysis of dental characters (Kimura, 2011, 2013), *S. prima* is recovered as the sister branch to all other fossil and recent *Sicista*, however, with just moderate support. This phylogenetic placement is consistent with the fact that the inferred time of basal radiation of crown *Sicista* (ca. 6 Mya) is much later than the age of *S. prima*.

After a gap spanning early Middle Miocene, the north Chinese record shows a presence of several lineages of birch mice. A long-living lineage of relatively simple-toothed *S. ertemteensis* ranges from late Middle Miocene to late Early Pliocene. Late Miocene to Early Pliocene, *S. bilikeensis* shows a slightly more complex dentition but is morphologically similar in general and may be conspecific to *S. ertemteensis*. Closely synchronous (Late Miocene to Early Pliocene, ca. 6.0–4.0 Mya), slightly larger *S. wangi* (Qiu & Li, 2016; Qiu & Storch, 2000) possibly indicates a radiation of the group. Another indication of a possible Late Miocene range expansion and radiation is the mid-Late Miocene record from Pavlodar (Kazakhstan) with *S. bagajevi* showing a complex-toothed dentition (Savinov, 1970). During the mid-Late Miocene, the group reached the Caucasus with *Sicista sp.* showing a morphologically primitive first lower molar (Tesakov et al., 2017). The emergence of multiple fossil lineages in the Late Miocene correlates well with nearly bush-like radiation among the five major recent lineages estimated to occur in the latest Miocene–earliest Pliocene.

However, it is yet unclear whether any of the fossil lineages are genealogically close to the recent ones. Dental morphology provides insufficient information as it is demonstrated by the low resolution of the trees inferred by Kimura (2013). Moreover, molar evolution is subject to parallelisms, which can obscure true phylogenetic signal. Thus, in the latter analysis, *S. bagajevi* is placed in the same clade as *S. betulina*, with which it shares complex tooth morphology, while *S. subtilis* stands as a deep branching lineage with no close relatives. This pattern can hardly be correct since all DNA sequence data strongly support close relationships between complex-toothed *S. betulina* group and simple-toothed *S. subtilis* group, therefore suggesting that the complex molar pattern evolved independently in ancient *S. bagajevi* and, more recently, in the ancestor of *S. betulina*.

Early Pliocene (5.3–3.5 Mya) birch mice are listed in localities of western Siberia (Zykin, 2012). In Late Pliocene (3.5–2.6 Mya), the oldest birch mice are known from the Transbaikal region as *S. pliocaenica* (Erbaeva, 1976). In early Early Pleistocene (Gelasian, 2.6–1.8 Mya), *Sicista* is listed from northeastern Kazakhstan (Zazhigin, 2009). In slightly younger faunas (ca. 2.0–1.8 Mya), *Sicista* specimens are reported from southern Kazakhstan (Tjutkova & Kaipova, 1996) and from southern West Siberia (Tesakov, Bondarev, & Frolov, 2016). Gelasian (2.6–1.8 Mya) *Sicista* from Europe are only known in sporadic localities of northern Black Sea region (Tesakov, 2004; Topachevsky, Scorik, & Rekovets, 1987). Starting from late Early Pleistocene (Calabrian, 1.8–0.8 Mya), the European fossil record shows multiple and geographically widespread remains of the group (Kowalski, 2001). Two fossil species, which are likely conspecific, were described from late Early Pleistocene faunas of Hungary, *Sicista praeloriger* (Kormos, 1930; Schaub, 1930), and southern Ukraine, *Sicista vinogradovi* (Topachevsky, 1965). Both forms occur in faunas with dominant forms indicating open landscapes. They have relatively large molars with a number of accessory elements intermediate between extant *S. subtilis* and *S. betulina* groups and may represent early stages of *S. subtilis*.

The first reliable records of the *S. subtilis* group in Europe date back to late Early and early Middle Pleistocene; the smaller and more complex-toothed *S. betulina* are known since mid-Middle Pleistocene (Kowalski, 2001). These data are in line with the results of molecular dating suggesting that the split between *S. betulina* and *S. subtilis* groups occurred at the Pliocene–Pleistocene boundary while the radiation within the groups started at the end of the Early Pleistocene.

In conclusion, several points should be outlined. The available fossil data suggest east Central Asia as the origin of the genus; this view is in apparent agreement with the molecular tree, given the basal position of the "*tianschanica*" and "*concolor*" lineages. The ancestral habitat of Late Miocene–earliest Pliocene *Sicista* is likely lowland mesic to semiarid steppe and forest-steppe as follows from the environmental polarities of the small mammal assemblages of well-studied faunas Ertemte and Bilike (Qiu, Wang, & Li, 2013).

Both fossil and molecular genetic data are consistent in suggesting that extensive radiation in *Sicista* began in the Late Miocene. It should be noted that eastern Central Asia was less affected by the Late Miocene aridization and retained many elements of mesic steppe fauna (Fortelius & Zhang, 2006). It may be hypothesized that an essential stage of the genus expansion is associated with the advancement of mesic forest-steppe across western Eurasia in the beginning of Pliocene (Vangengeim, Vislobokova, & Sotnikova, 1998). The position of *S. caudata* as the sister group to the *S. caucasica* species group suggests that there could have been several waves of colonization of large areas of Eastern Europe and Northern Asia by *Sicista*. The common ancestor of these two lineages may have dispersed over the territory from the Far East to the Caucasus during the Late Miocene–Pliocene transition period but later disappeared from Central Asia, Siberia, and Eastern Europe. The latter event could have been a result of the climate change at the Pliocene–Pleistocene transition and may be associated with the replacement by the ancestor of the *S. subtilis* group, which could be better adapted to arid environment.

The fact that many birch mouse species are now restricted to mountain areas does not imply mountainous origin of the genus as it was argued in Cserkész et al. (2019). Based on fossil record and paleogeographic reconstructions, one may hypothesize that the contemporary distribution of *S. concolor*, *S. tianschanica*, *S. napaea*, and *S. caucasica* is rather a result of several independent parallel habitat shifts to higher altitudes in response to Plio-Pleistocene aridization. It may be suggested that such shifts became possible because birch mice were preadapted to cold and high-altitude habitats as they are specialized hibernators. The latter feature is probably a plesiomorphy shared by all Dipodoidea, most of which hibernate during the cold season. In contrast to dipodoids, its specious sister group Muroidea includes just a few hibernating species (some Cricetinae), although many muroids inhabit the same environment as Sminthidae and Dipodidae.

ACKNOWLEDGEMENTS

The research was supported by the Russian Foundation for Basic Research (grant 18-04-00400a for VM & VL; 17-04-00065a for AB & VL). Phylogenetic analysis was performed by VL within the framework of the Russian Science Foundation, project No. 14-50-00029. Special thanks to Ya. Red'kin, A. Surov, and I. Tikhonov for donating samples of birch mice. The authors would like to thank the three anonymous referees and the editor for their valuable comments.

ORCID

Vladimir S. Lebedev https://orcid.org/0000-0003-4943-7097
Mikhail Yu. Rusin https://orcid.org/0000-0003-4349-4795
Elena D. Zemlemerova https://orcid.org/0000-0002-3624-853X
Vera A. Matrosova https://orcid.org/0000-0003-2325-8584
Alexey S. Tesakov https://orcid.org/0000-0002-8616-2291

REFERENCES

Bannikova, A. A., Lebedev, V. S., Abramov, A. V., & Rozhnov, V. V. (2014). Contrasting evolutionary history of hedgehogs and gymnures (Mammalia: Erinaceomorpha) as inferred from a multigene study. *Biological Journal of the Linnean Society*, *112*, 499–519 https://doi.org/10.1111/bij.12299

Baskevich, M. I. (2016). Taxonomy, evolution and variation of the genus *Sicista* (Rodentia, Dipodoidea): A review of karyological and molecular data. *Archives of Zoological Museum of Lomonosov Moscow State University*, *54*, 191–228 (in Russian).

Baskevich, M. I., Okulova, N. M., Oparin, M. L., & Vlasov, A. A. (2005). *Chromosomal and craniometric variability in Sicista strandi (Rodentia, Dipodoidea) from the Caucasus and Russian plain*. In Mammals of mountain areas. Proceedings of the International Conference (p. 18–22). Moscow, KMK Press (in Russian).

Baskevich, M. I., Potapov, S. G., Okulova, N. M., Khlyap, L.A., Oparin, M. L., & Shvarts, E. A., … Stakheev, V. V. (2015). Species structure in birch mice of the "*betulina*" group (Rodenta, Dipodoidea, *Sicista*) of East Europe in the light of chromosome, molecular, and craniometric data. In *Species structure in mammals. Materials of the conference* (p.14). *Moscow.* KMK Press. (in Russian).

Baum, B. R. (1992). Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon*, *41*, 3–10 https://doi.org/10.2307/1222480

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bobrinsky, N. A., Kuznetsov, B. A., & Kuzyakin, A. P. (1965). *Opredelitl' mlekopitayushchikh SSSR [Guide to the mammals of the U.S.S.R.]* (2nd ed.). Moscow, Russia: Prosveshchenie. (in Russian).

Bradley, R. D., & Baker, R. J. (2001). A test of the genetic species concept: Cytochrome-*b* sequences and mammals. *Journal of Mammalogy*, *82*, 960–973. https://doi.org/10.1644/1545-1542(2001)082<0960:ATOTGS>2.0.CO;2

Corbet, G. B. (1978). *The mammals of the Palaearctic region: A taxonomic review*. London, UK: British Museum (Natural History).

Cserkész, T., Fülöp, A., Almerekova, S., Kondor, T., Laczkó, L., & Sramkó, G. (2019). Phylogenetic and morphological analysis of birch mice (Genus *Sicista*, Family Sminthidae, Rodentia) in the Kazak Cradle with description of a new species. *Journal of Mammalian Evolution*, *26*, 147–163. https://doi.org/10.1007/s10914-017-9409-6

Cserkész, T., Rusin, M., & Sramkó, G. (2016). An integrative systematic revision of the European southern birch mice (Rodentia: Sminthidae, *Sicista subtilis* group). *Mammal Review*, *46*, 114–130. https://doi.org/10.1111/mam.12058

Deng, C., Hao, Q., Guo, Z., & Zhu, R. (2019). Quaternary integrative stratigraphy and timescale of China. *Science China Earth Sciences*, *62*, 324–348. https://doi.org/10.1007/s11430-017-9195-4

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*, 1969–1973. https://doi.org/10.1093/molbev/mss075

Ellerman, J. R., & Morrison-Scott, T. C. S. (1951). *Checklist of Palaearctic and Indian mammals 1758 to 1946*. London, UK: Trustees of the British Museum (Natural History).

Erbaeva, M. A. (1976). Fossil bunodont rodents in Transbaikal area. *Geologiya i Geofizika*, *17*, 144–149. (in Russian).

Fortelius, M., & Zhang, Z. (2006). An oasis in the desert? History of endemism and climate in the late Neogene of North China. *Palaeontographica Abteilung A*, *277*, 131–141.

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Proceedings*

of the Conference held in Nucleic acids symposium series, *41*, 95–98.

Heled, J., & Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, *27*, 570–580. https://doi.org/10.1093/molbev/msp274

Holden, M. E., Cserkész, T., & Musser, G. (2017). Family Sminthidae. In D. E. Wilson, T. E. Lacher Jr & R. A. Mittermeier (Eds.) *Handbook of the Mammals of the World*, *Volume 7*, *Rodents II* (pp. 22–48). Barcelona, Spain: Lynx Ediciones.

Holden, M. E., & Musser, G. G. (2005). Superfamily Dipodoidea. In D. E. Wilson & D. M. Reeder (Eds.), *Mammal species of the world*, 3rd ed. (pp. 871–893). Baltimore, MD: Johns Hopkins University Press.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*, 587–589. https://doi.org/10.1038/nmeth.4285

Kimura, Y. (2011). The earliest record of birch mice from the Early Miocene Nei Mongol, China. *Naturwissenschaften*, *98*, 87–95. https://doi.org/10.1007/s00114-010-0744-1

Kimura, Y. (2013). Intercontinental Dispersals of Sicistine Rodents (Sicistinae, Dipodidae, Rodentia) Between Eurasia and North America. In X. Wang, L. J. Flynn & M. Forterius (Eds.), *Fossil Mammals of Asia: Neogene biostratigraphy and chronology* (pp. 655–675). New York, NY: Columbia University Press.

Kodandaramaiah, U. (2010). Use of dispersal–vicariance analysis in biogeography – a critique. *Journal of Biogeography*, *37*, 3–11.

Kormos, T. (1930). Diagnosen neuer Säugetiere aus der ober-Pliozänen Fauna des Somlyoberges bei Püspökfürdő. *Annales Musei Nationalis Hungarici*, *27*, 237–246 (in German).

Kovalskaya, Y. M., Aniskin, V. M., Bogomolov, P. L., Surov, A. V., Tikhonov, I. A., Tikhonova, G. N., … Volobouev, V. T. (2011). Karyotype reorganisation in the subtilis group of birch mice (Rodentia, Dipodidae, *Sicista*): Unexpected taxonomic diversity within a limited distribution. *Cytogenetic and Genome Research*, *132*, 271–288. https://doi.org/10.1159/000322823

Kowalski, K. (2001). Pleistocene rodents of Europe. *Folia Quaternaria*, *72*, 1–382.

Lebedev, V. S., Bannikova, A. A., Pagès, M., Pisano, J., Michaux, J. R., & Shenbrot, G. I. (2013). Molecular phylogeny and systematics of Dipodoidea: A test of morphology-based hypotheses. *Zoologica Scripta*, *42*, 231–249. https://doi.org/10.1111/zsc.12002

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*, 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Matzke, N. J. (2013). BioGeoBEARS: biogeography with Bayesian (and likelihood) evolutionary analysis in R scripts. *R package*, *version 0.2*, *1*, 2013.

Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, *30*, 1188–1195. https://doi.org/10.1093/molbev/mst024

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution*, *32*, 268–274. https://doi.org/10.1093/molbev/msu300

Ognev, S. I. (1948). *Mammals of USSR and adjacent countries. (The mammals of Eastern Europe and Northern Asia). Volume 6*. Moscow & Leningrad, Russia: Izdatelstvo Akademii Nauk SSSR(In Russian).

Pirie, M. D., Humphreys, A. M., Antonelli, A., Galley, C., & Linder, H. P. (2012). Model uncertainty in ancestral area reconstruction: A parsimonious solution? *Taxon*, *61*, 652–664. https://doi.org/10.1002/tax.613013

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). *Journal of Biogeography*, *42*, 856–870. https://doi.org/10.1111/jbi.12476

Qiu, Z.. (2017). Chapter 7. Yushe Dipodoidea. In L. Flynn & W.-Y. Wu (Eds.), *Late Cenozoic Yushe Basin*, *Shanxi Province*, *China: Geology and fossil mammals. Volume II: Small mammal fossils of Yushe Basin* (pp. 81–86). Dordrecht, Netherlands: Springer Science+Business Media. https://doi.org/10.1007/978-94-024-1050-1

Qiu, Zh, & Li, Q. (2016). Neogene rodents from central Nei Mongol, China. *Palaeontologia Sinica*, *New Series C*, *30*, 1–684.

Qiu, Zh, & Storch, G. (2000). The early Pliocene micromammalian fauna of bilike, Inner Mongolia, China (Mammalia: Lipotyphla, Chiroptera, Rodentia, Lagomorpha). *Senckenbergiana Lethaea*, *80*, 173–229. https://doi.org/10.1007/BF03043669

Qiu, Z. D., Wang, X., & Li, Q. (2013). Neogene faunal succession and biochronology of central Nei Mongol (Inner Mongolia). In X. Wang, L. J. Flynn & M. Fortelius (Eds.), *Fossil mammals of Asia: Neogene biostratigraphy and chronology* (pp. 155–186). New York, NY: Columbia University Press. https://doi.org/10.7312/columbia/9780231150125.001.0001

Ragan, M. A. (1992). Phylogenetic inference based on matrix representation of trees. *Molecular Phylogenetics and Evolution*, *1*, 53–58. https://doi.org/10.1016/1055-7903(92)90035-F

Rambaut, A., & Drummond, A. (2003). Tracer. A program for analysing results from Bayesian MCMC programs such as BEAST & MrBayes. *University of Edinburgh*, *UK*. Version 1.6. http://beast.bio.ed.ac.uk/

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., … Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*, 539–542. https://doi.org/10.1093/sysbio/sys029

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, the Italian Journal of Mammalogy*, *29*, 61–66. https://doi.org/10.4404/hystrix-00050-2018

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: A laboratory manual* (2nd ed.). New York, NY: Cold spring harbor laboratory press.

Savinov, P. F. (1970). Jerboas (Dipodidae, Rodentia) from the Neogene of Kazakhstan. In K. K. Flerov (Ed.), *Material on evolution of terrestrial vertebrates* (pp. 91–134). Moscow, Russia: Nauka. (in Russian).

Schaub, S. (1930). Fossile Sicistinae. *Eclogae Geologicae Helvetiae*, *23*, 616–637.

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: Dipodinae) from China: A living fossil? *Journal of Zoological Systematics and Evolutionary Research*, *55*, 356–368. https://doi.org/10.1111/jzs.12182

Shenbrot, G. I., Sokolov, V. E., Geptner, V. G., & Kovalskaya, Y. M. (1995). *Mammals of Russia and adjacent regions. Dipodoidae*. Moscow, Russia: Nauka. (in Russin).

Sokolov, V. E., Baskevich, M. I., & Kovalskaya, Y. M. (1981). Revision of birch mice of the Caucasus: Sibling species *Sicista caucasica* Vinogradov, 1925 and *S. klucorica* sp.n. (Rodentia, Dipodidae). *Zoologicheskii Zhurnal*, *60*, 1386–1393 (in Russian).

Sokolov, V. E., & Kovalskaya, Y. M. (1990). Karyotypes of birch mice (*Sicista*, Dipodoidea, Rodentia) in the northern Tien-Shan and Sikhote-Alin. *Zoologicheskii Zhurnal*, *69*, 152–157 (in Russian).

Sokolov, V. E., Kovalskaya, Y. M., & Baskevich, M. I. (1982). Taxonomy and comparative cytogenetics of some species of the genus *Sicista* (Rodentia, Dipodidae). *Zoologicheskii Zhurnal*, *61*, 102–108 (in Russian).

Sokolov, V. E., Kovalskaya, Y. M., & Baskevich, M. I. (1987). Review of karyological research and the problems of systematics in the genus *Sicista* (Zapodidae, Rodentia, Mammalia). *Folia Zoologica*, *36*, 35–44.

Sokolov, V. E., Kovalskaya, Y. M., & Baskevich, M. I. (1989). On species status of northern birch mice *Sicista strandi* (Rodentia, Dipodidae). *Zoologicheskii Zhurnal*, *68*, 95–106 (in Russian).

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics*, *68*, 978–989. https://doi.org/10.1086/319501

Swofford, D. L. (2003). PAUP*: phylogenetic analysis using parsimony, version 4.0 b10.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA 6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*, 2725–2729. https://doi.org/10.1093/molbev/mst197

Tesakov, A. S. (2004). *Biostratigraphy of Middle Pliocene – Eopleistocene of Eastern Europe (based on small mammals)*. Moscow, Russia: Nauka. (in Russian).

Tesakov, A. S., Bondarev, A. A., & Frolov, P. D. (2016). Isakovka 4, a new locality of Early Pleistocene terrestrial biota in West Siberia. In V. V. Arkadiev et al. (Eds.), *Materials of the LXII session of the Paleontological Society, April 4-6* (pp. 276–277). St.-Petersburg: Russian Geological Research Institute. (in Russian).

Tesakov, A. S., Titov, V. V., Simakova, A. N., Frolov, P. D., Syromyatnikova, E. V., Kurshakov, S. V., … Zelenkov, N. V. (2017). Late Miocene (Early Turolian) vertebrate faunas and associated biotic record of the Northern Caucasus: Geology, palaeoenvironment, biochronology. *Fossil Imprint*, *73*, 383–444. https://doi.org/10.2478/if-2017-0021

Tjutkova, L. A., & Kaipova, G. O. (1996). Late Pliocene and Eopleistocene micromammal faunas of southeastern Kazakhstan. *Acta Zoologica Cracoviensia*, *39*, 549–557.

Topachevsky, V. A. (1965). *Insectivores and rodents of the Late Pliocene fauna of Nogaisk*. Kiev, Ukraine: Naukova Dumka. (in Russian).

Topachevsky, V. A., Scorik, A. F., & Rekovets, L. I. (1987). *Rodents of the Upper Neogene and Early Anthropogene deposits of the Khadjibei Lagoon*. Kiev, Ukraine: Naukova Dumka. (in Russian).

Vangengeim, E. A., Vislobokova, I. A., & Sotnikova, M. V. (1998). Large Ruscinian Mammalia in the territory of the former Soviet Union. *Stratigraphy and Geological Correlation*, *6*, 368–382.

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088

Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, *107*, 9264–9269. https://doi.org/10.1073/pnas.0913022107

Yu, Y., Harris, A. J., Blair, C., & He, X. (2015). RASP (Reconstruct Ancestral State in Phylogenies): A tool for historical biogeography. *Molecular Phylogenetics and Evolution*, *87*, 46–49. https://doi.org/10.1016/j.ympev.2015.03.008

Zazhigin, V. S. (2009). Complexes of the small mammals of the Late Pliocene - Early Pleistocene of the south of the West Siberia. In A. E. Kontorovich et al. (Eds.), *Fundamental problems of the Quaternary: Results of the studies and main directions of future research. Material of the 6th All-Russia Conference of the Study of Quaternary, November 19–23, 2009* (pp. 218–220). Novosibirsk, Russia: Siberian Branch of the Russian Academy of Sciences. (in Russian).

Zhang, Y., Jin, S., Quan, G., Li, S., Ye, Z., Wang, F., & Zhang, M. (1997). *Distribution of mammalian species in China*. Beijing, China: China Forestry Publishing House.

Zykin, V. S. (2012). *Stratigraphy and evolution of environments and climate during late cenozoic in the southern West Siberia*. Novosibirsk, Russia: Geo. (in Russian).

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Lebedev VS, Rusin MY, Zemlemerova ED, et al. Phylogeny and evolutionary history of birch mice *Sicista* Griffith, 1827 (Sminthidae, Rodentia): Implications from a multigene study. *J Zool Syst Evol Res*. 2019;57:695–709. https://doi.org/10.1111/jzs.12279

APPENDIX A

Information on the material analyzed: species, ID of specimens, collection numbers of voucher specimens as well as geographic origin. GenBank accession numbers are given in Supporting information Table S1.

Species	ID of specimen	Voucher	Geographic location
Sicista caudata	7	ZMMU S-188665	Russia, Sakhalin Island, Pilenga R.
S. caudata	17	ZMMU S-184768	Russia, Sakhalin Island, Listvenitsa R.
S. armenica	2015		Armenia, Gegharkunik, Semenovka
S. kazbegica	1		Russia, North Ossetia, Unal
S. kazbegica	2		Russia, North Ossetia, Unal
S. kazbegica	T0761	T0761	Georgia, Kazbegi, Suatisi
S. kazbegica	T0762	T0762	Russia, North Ossetia, Tsey
S. kazbegica	Tsey 2		Russia, North Ossetia, Tsey
S. kazbegica	Tsey 3		Russia, North Ossetia, Tsey
S. caucasica	1	ZMMU S-196002	Russia, Adygea, Lagonaki
S. caucasica	2	ZMMU S-196003	Russia, Adygea, Lagonaki
S. caucasica	3	ZMMU S-196004	Russia, Adygea, Lagonaki
S. caucasica	T0764	T0764	Russia, West Caucasus
S. caucasica	Adler		Russia, Krasnodar, Mzymta
S. kluchorica	A	ZMMU S-197464	Russia, Karachay-Cherkessia, Teberda
S. kluchorica	B	ZMMU S-197465	Russia, Karachay-Cherkessia, Teberda
S. kluchorica	C	ZMMU S-197466	Russia, Karachay-Cherkessia, Teberda

(Continues)

APPENDIX (Continued)

Species	ID of specimen	Voucher	Geographic location
S. kluchorica	D	ZMMU S-197467	Russia, Karachay-Cherkessia, Teberda
S. kluchorica	T0763	T0763	Russia, Kabardino-Balkaria, Elbrus
S. kluchorica	Elbrus 2		Russia, Kabardino-Balkaria, Elbrus
S. kluchorica	Elbrus 3		Russia, Kabardino-Balkaria, Elbrus
S. kluchorica	Elbrus 4		Russia, Kabardino-Balkaria, Elbrus
S. pseudonapaea	T-MNHN1999-451	T-MNHN1999-451	Altai
S. pseudonapaea	T-MNHN1999-452	T-MNHN1999-452	Altai
S. betulina	17		Russia, Moscow reg., Chernogolovka
S. betulina	15.1		Russia, Altai, Artybash
S. strandi	1	ZMMU S-181441	Russia, Belgorod reg., Gubkinskiy distr.
S. strandi	21		Russia, Belgorod reg., Gubkinskiy distr.
S. strandi	14	ZMMU S-178460	Russia, Rostov reg., Tsimla sands
S. strandi	15_1	ZMKNU 7542	Ukraine, Lugansk reg., Provalye
S. strandi	30		Russia, Saratov reg., Slavianka
S. nordmanni	26		Russia, Belgorod reg., Borisovka
S. nordmanni	SSN01	ZMKNU 7541	Ukraine, Kherson reg., Geroysk
S. nordmanni	3		Russia, Belgorod reg., Borisovka
S. subtilis	32		Russia, Tuva
S. subtilis	18	ZMMU S-182802	Kazakhstan, Pavlodar reg., Kudaykol
S. subtilis	v16_113	ZMMU S-197172	Russia, Astrakhan reg., near Grachi
S. subtilis	11		Russia, Volgograd reg., Kamyshin dis.
S. concolor	MK0509BL02	MK0509BL02	China, Sichuan prov., Maerkang
S. concolor	KJ648496		China, Sichuan prov., Baoxing county
S. tianschanica	Trb3	ISEA 59532	Kazakhstan, East Kazakhstan, Urzhar
S. tianschanica	NT0609AQ10	NT0609AQ10	China, Xinjiang, Narati
S. tianschanica	NT0609BR03	NT0609BR03	China, Xinjiang, Narati
S. cimlanica	SS10		Russia, Tsimla sands
S. cimlanica	5		Russia, Tsimla sands
S. trizona	SSU65		Romania, Feiurdeni
S. trizona	SSU56		Hungary, Mezocsat
Paralactaga elater		ZMMU S-198794 (original sequences)	Uzbekistan, Navoi prov.
Cardiocranius paradoxus		ZMMU S-188976 (original sequences)	Mongolia, Govi-Altai prov.
Eozapus setchuanus		Data taken from GenBank	
Napaeozapus insignis		Data taken from GenBank	
Zapus princeps		Data taken from GenBank	
Zapus_trinotatus		Data taken from GenBank	
Eozapus setchuanus		Data taken from GenBank	
Dipus sagitta		Data taken from GenBank	
Jaculus jaculus		Data taken from GenBank	

ZMMU, Zoological Museum of Moscow University.