

RESEARCH ARTICLE

Implications of Hybridization, NUMTs, and Overlooked Diversity for DNA Barcoding of Eurasian Ground Squirrels

Oleg A. Ermakov^{1*‡}, Evgeniy Simonov^{2,3‡}, Vadim L. Surin⁴, Sergey V. Titov¹, Oleg V. Brandler⁵, Natalia V. Ivanova⁶, Alex V. Borisenko⁶

1 Department of Zoology and Ecology, Penza State University, Penza, Russia, **2** Institute of Systematics and Ecology of Animals, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia, **3** Tomsk State University, Tomsk, Russia, **4** Hematological Research Center, Russian Academy of Medical Sciences, Moscow, Russia, **5** Koltzov Institute of Developmental Biology, Russian Academy of Sciences, Moscow, Russia, **6** Biodiversity Institute of Ontario, University of Guelph, Guelph, Canada

‡ These authors contributed equally to this work.

* oaermakov@list.ru



Abstract

The utility of DNA Barcoding for species identification and discovery has catalyzed a concerted effort to build the global reference library; however, many animal groups of economical or conservational importance remain poorly represented. This study aims to contribute DNA barcode records for all ground squirrel species (Xerinae, Sciuridae, Rodentia) inhabiting Eurasia and to test efficiency of this approach for species discrimination. Cytochrome c oxidase subunit 1 (COI) gene sequences were obtained for 97 individuals representing 16 ground squirrel species of which 12 were correctly identified. Taxonomic allocation of some specimens within four species was complicated by geographically restricted mtDNA introgression. Exclusion of individuals with introgressed mtDNA allowed reaching a 91.6% identification success rate. Significant COI divergence (3.5–4.4%) was observed within the most widespread ground squirrel species (*Spermophilus erythrogenys*, *S. pygmaeus*, *S. suslicus*, *Urocitellus undulatus*), suggesting the presence of cryptic species. A single putative NUMT (nuclear mitochondrial pseudogene) sequence was recovered during molecular analysis; mitochondrial COI from this sample was amplified following re-extraction of DNA. Our data show high discrimination ability of 100 bp COI fragments for Eurasian ground squirrels (84.3%) with no incorrect assessments, underscoring the potential utility of the existing reference library for the development of diagnostic ‘mini-barcodes’.

OPEN ACCESS

Citation: Ermakov OA, Simonov E, Surin VL, Titov SV, Brandler OV, Ivanova NV, et al. (2015) Implications of Hybridization, NUMTs, and Overlooked Diversity for DNA Barcoding of Eurasian Ground Squirrels. PLoS ONE 10(1): e0117201. doi:10.1371/journal.pone.0117201

Academic Editor: Laura M. Boykin, The University of Western Australia, AUSTRALIA

Received: September 17, 2014

Accepted: December 19, 2014

Published: January 24, 2015

Copyright: © 2015 Ermakov et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The full dataset containing sequencing trace files and sequences is available as the BOLD dataset “Ground Squirrels of the Palaearctic – Comparative Dataset” (DS-GSPA), DOI: [dx.doi.org/10.5883/DS-GSPA](https://doi.org/10.5883/DS-GSPA)

Introduction

DNA barcoding [1] has proved to be a useful tool for species identification (e.g. [2–5]) and serving for various needs from forensic analysis (e.g. [6]) to biodiversity surveys (e.g. [7]). The 5-prime ~650 base pair region of the mitochondrial cytochrome oxidase subunit I gene (COI) is the generally accepted standard DNA barcode marker for most species of animals. Some

applications, like noninvasive analysis of old samples (e.g. [8]) or the emerging DNA metabarcoding [9], require shorter markers (~50–200 bp). An important challenge to this approach is posed by the introgression of mitochondrial DNA due to hybridization and/or incomplete lineage sorting of mtDNA haplotypes (e.g. [10]). Both of them can lead to the absence of the “barcoding gap” and cause misidentification. An essential prerequisite in the utility of DNA barcoding for practical applications is the creation of a high-quality reference database that is scrutinized for possible analytical and taxonomic errors. Despite a concerted effort to attain broad representation of key taxa in the reference library hosted by the Barcode of Life Data System (BOLD—www.boldsystems.org) many groups of economical or conservational importance remain poorly represented.

Ground squirrels (Marmotini) are a charismatic faunal element of grassland communities across north-temperate biomes of the Holarctic and play an important role in maintaining these open habitats. In Eurasia, this group is represented by 16 species belonging to the genera *Spermophilus* and *Urocitellus* [11]. *Spermophilus* has exclusively Palaearctic distribution, while *Urocitellus* is predominantly Nearctic, with only two species occurring in Easternmost Siberia. The native range of Eurasian ground squirrels spans a vast area from Central Europe and the Middle East to the Chukotka Peninsula [12, 13]. Historically, these animals have been regarded as major agriculture pests [14]. In addition, they were found to be important reservoirs of dangerous natural-focal zoonotic infections, such as plague, rabbit-fever, relapsing fever, Q fever, brucellosis, etc. [15–17]. Triggered by these findings, concerted eradication efforts have been deployed across of Eurasia throughout much of the XX century [18]. Coupled with extensive agricultural transformation of grassland habitats, this has led to significant population decline and range fragmentation in many ground squirrel species (e.g. [19, 20]).

Today, the trend has shifted from extermination to protection, which is manifested by a growing number of ground squirrel conservation and reintroduction programs in parts of Europe (e.g. [21, 22]). Several Eurasian ground squirrels have special global conservation status in the IUCN Red List: three species (*S. musicus*, *S. suslicus* and *S. xanthoprymnus*) are listed as ‘Near Threatened’ and one (*S. citellus*) – as ‘Vulnerable’ [23]. Many Central and East European populations are facing local threats from human environmental impact and are included in national and regional Red Data Lists; some of them are presently believed to be extinct [24]. For instance, the Red-cheeked ground squirrel (*S. erythrogegnys*) is now extinct from its type locality (our data). Assessment of conservation status in some ground squirrels is hampered by ongoing debates about the taxonomic rank of certain named forms [25, 26]. The recent description of a new species from Turkey – *S. taurensis* [27, 28] suggests that taxonomic knowledge gaps remain even within this relatively well studied group of mammals. The existence of unresolved systematic questions, combined with conservational and epizootological importance of Eurasian ground squirrels calls for continued taxonomic reassessments employing novel methodological approaches and for the development of new diagnostic tools.

This study aims to establish the COI barcode reference library for all ground squirrel species inhabiting Eurasia, to assess its utility for species discrimination, to highlight any previously unrecognised genetic diversity, and to discuss possible implications of mitochondrial introgression.

Materials and Methods

Sample collection

The studied material represents all 16 presently recognized ground squirrel species from Eurasia (genera *Spermophilus* and *Urocitellus*) and includes all species from the former genus (Fig. 1). Geographic sampling spans 74 different locations. No experiments were conducted

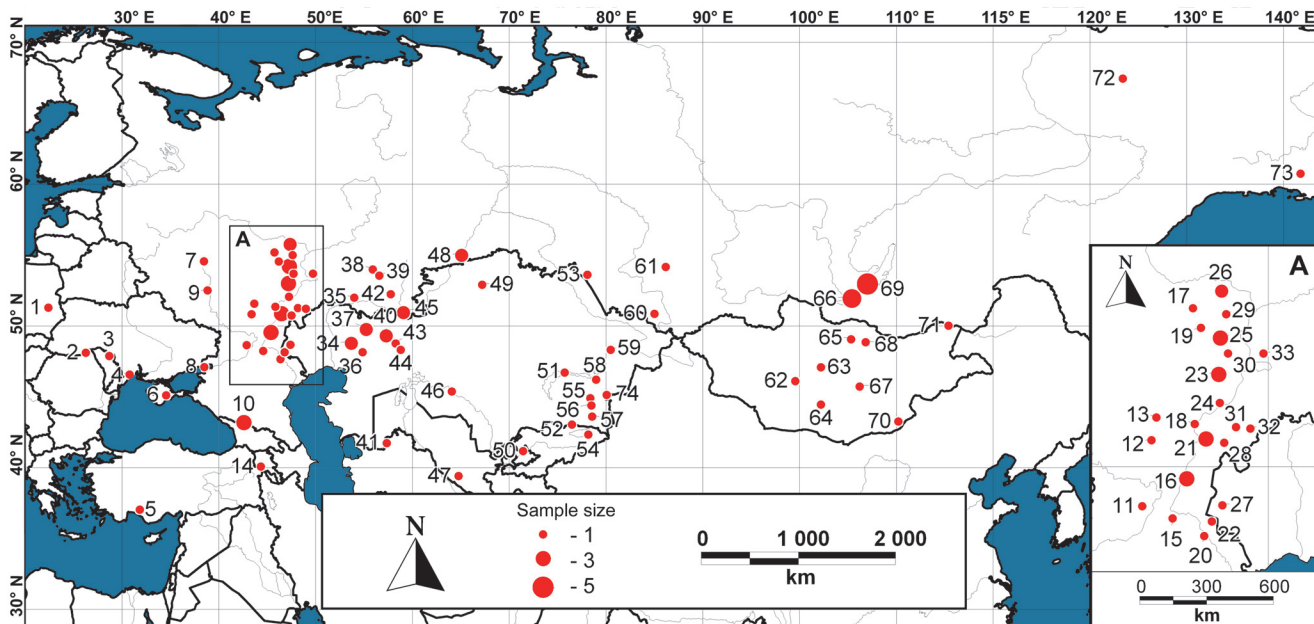


Figure 1. Map of the study area showing collection localities for this study (see S1 Table for locality information).

doi:10.1371/journal.pone.0117201.g001

with living animals. All ground squirrel and “outgroup” samples came from preserved tissue from vouchered collection specimens deposited in the following institutions: Penza State University (PSPU; 69 samples); Koltzov Institute of Developmental Biology, Russian Academy of Sciences (IDB; 14 samples); Zoological Museum of Moscow State University (ZMMU; 11 samples); Zoological Museum, Institute of Systematics and Ecology of Animals, Siberian Branch, Russian Academy of Sciences (ISEA; five samples); Charles University in Prague (CU; one sample); Zoological Institute, Russian Academy of Sciences (ZISP; one sample). Museum catalogue numbers along with repository and locality data are given in S1 Table. All specimens used in this study were morphologically identified prior to sequencing; taxonomy follows Helgen *et al.* [11].

DNA isolation, amplification and sequencing

Extraction of total DNA and subsequent analyses was done either at the Biodiversity Institute of Ontario, Guelph, following protocols provided in Ivanova *et al.* [29], or at the Laboratory of Animal Systematics and Molecular Ecology, Penza State University, following the protocols described below. DNA was extracted according to a standard procedure including the treatment with sodium dodecyl sulphate and proteinase K, and subsequent phenol-chloroform extraction [30]. To deal with degraded DNA from old museum samples, we developed two primer pairs specific to ground squirrels (subfamily Marmotinae): Sp-COXD – 5'-GAT GAT TCT TCT CAA CTA ATC-3' and SpCOXRr – 5'-CAT GGG CRA GAT TTC CAG CTA-3'; SpCOXDd – 5'-CTT CTA TRG TTG AAG CAG GTG C-3' and SpCOXR – 5'-TGA GAA ATT ATA CCA AAT CCT G-3'. Each PCR reaction contained 50 mM Tris-HCl (pH 8.9), 20 mM ammonium sulphate, 20 μM EDTA, 150 μg/ml bovine serum albumin, dNTPs (200 μM of each), 2 mM MgCl₂, 15 pmol of each primer, 2 units of *Taq* polymerase and 0.1 to 0.2 μg DNA in a final volume of 25 μl. The reaction conditions were 94°C for 1 min; 62°C for 1 min; and 72°C for 1 min (30 cycles). PCR products were analysed using electrophoresis in 6% PAAG with subsequent staining with ethidium bromide and visualization in the UV light. Sequencing was done on an

ABI 3500 automated capillary sequencer (Applied Biosystems) with the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit 3.1 using the same primers. Sequences were aligned manually and checked for unexpected stop codons using BioEdit 7.0 [31].

Data analysis

To complement our analysis, additional sequences of New World ground squirrels and other selected members of the family Sciuridae were obtained from the BOLD project “Mammals of Canada” (ABMC) and from GenBank (S1 Table). Although DNA barcoding is not a phylogenetic approach, we used MetaPIGA2 [32] to infer the gene tree using Maximum Likelihood (ML) that was compared against the branching pattern inferred from the ‘conventional’ Neighbour-Joining (NJ) method. Before running ML analysis, the dataset was tested for redundancy and transition saturation using the same program. The default substitution model used by BOLD (www.boldsystems.org) is K2P model [33]; however the use of this model in DNA-barcoding has been criticized (i.e. [34, 35]). Thus, we determined the best-fitting models of nucleotide substitution for our data using jModelTest 2.1.1 [36] with Akaike Information Criterion (AIC). We used a variable number of bootstrap replicates, stopping the iterations when the mean relative error among 10 consecutive consensus trees stayed below 5% (minimum 100, maximum 10 000). Tree topologies resulting from ML analyses were visualised and edited using FigTree v.1.4 [37].

Intra- and interspecies genetic distances (p -distances) and their standard errors based on 1000 bootstrap replications were calculated using MEGA 5.1 [38]. The number of haplotypes, haplotype diversity and nucleotide diversity per site for each species were computed with the help of DnaSP v. 5.10.1 software [39]. To test barcoding efficiency in our dataset, we used the DNA barcoding package Spider v. 1.1–5 [40] for R [41]. Two methods which mimic the p -distance-based “species identification” algorithm used by BOLD were applied: the “threshID” function in Spider and the “best close match” criterion of Meier *et al.* [42]. Two approaches were used in threshold selection: the first one included optimisation procedure which minimises false-positive and false-negative errors for a range of threshold values (0.1–4% in 0.1% increments); the second approach is the experimental method implemented in Spider (“localMinima” function) which produces a density object from the distance matrix and determines where a dip in the density of genetic distances indicates the transition between intra- and inter-specific distances. Species represented by only one individual (singletons) were excluded from these analyses. Additionally, we checked the dataset for the presence of the “barcoding gap” [43] by calculating the furthest intraspecific distance and the closest non-conspecific distance for each individual in the dataset [44] and by producing its graphical representation.

In addition, we applied two recently developed methods for automated species delineation to explore possible cryptic diversity in Eurasian ground squirrels. Bayesian implementation of the Poisson tree processes (PTP) model [45] was tested to delimit species on the ML tree generated by MetaPIGA2 (without outgroups), using the bPTP server (<http://species.h-its.org>) with the following settings: 200 000 MCMC generations; thinning interval of 100 and first 15% was discarded as burn-in. Analysis were run three times with different random seed to ensure consistency of results between runs; a convergence within each run was assessed by examining of a likelihood trace plot. The second approach used was the refined single linkage (RESL) analysis, introduced by Ratnasingham and Hebert in 2013 [46] along with the Barcode Index Number (BIN) system. RESL is a multi-step process serving to assign DNA barcode sequences to operational taxonomic units (OTUs). Then, each OTU is assigned to a uniform resource identifier within the BIN system [46]. These steps were run on BOLD (<http://www.boldsystems.org>) as part of its operational routine and include all COI records stored in its reference library.

In view of the rapid advent of DNA metabarcoding studies, we examined our dataset for the regions potentially useful for generation of mini-barcodes (50–100 bp long) via the slideBox-plots function implemented in Spider. The distribution of pairwise genetic distances of each window was calculated and plotted for 50-bp and 100-bp width windows using a 3-bp (codon) interval. At the next step, a number of 50- and 100-bp windows with highest divergence were chosen for threshold value optimization and tested for barcoding efficiency as described above.

The data used in this study (sequences, trace files, and associated detailed specimen information) are available online at <http://www.boldsystems.org> in the published BOLD project “Ground Squirrels of the Palaearctic” (ABGPA). The full dataset containing these records and other published records used for comparison is available as the BOLD dataset “Ground Squirrels of the Palaearctic – Comparative Dataset” (DS-GSPA), DOI: [dx.doi.org/10.5883/DS-GSPA](https://doi.org/10.5883/DS-GSPA). Original sequence data were also deposited in NCBI GenBank, accessions KM537885 – KM537985 (S1 Table).

Results

COI barcode sequences (657 bp long) were obtained for 97 individuals of 16 ground squirrel species from 74 locations (Fig 1). Thus, all ground squirrel species known from Eurasia were covered for the first time. The alignment contained 188 variable positions, of which 166 were parsimony-informative. The mean transition/transversion ratio (over all sequence pairs) was 5.741, and the mean base composition was A: 26.4, C: 23.6, G: 16.2 and T: 33.8%. The mean *p*-distances between species within genera *Spermophilus* and *Urocitellus* were 6.9 and 11.3%, respectively. The mean distance to the nearest neighbour species within all Eurasian ground squirrels ranged from 0.5% (*S. brevicauda* and *S. major*) to 7.9% (*S. xanthoprimum* and *S. taurensis*), with a mean of 4.4% (Table 1). Maximum intraspecific distances were observed

Table 1. Intraspecific genetic variation of Eurasian Ground Squirrels.

Species name	n/N	<i>h</i> ± SD	π ± SD	Mean intra-sp ± SE	Max intra-sp	Nearest species	Distance to NS ± SE
<i>Spermophilus alashanicus</i>	2/2	1.00 ± 0.50	0.002 ± 0.001	0.002 ± 0.002	0.002	<i>S. citellus</i>	0.073 ± 0.010
<i>Spermophilus brevicauda</i>	4/3	0.83 ± 0.22	0.002 ± 0.001	0.002 ± 0.001	0.005	<i>S. major</i>	0.005 ± 0.002
<i>Spermophilus citellus</i>	1/1	-	-	-	-	<i>S. alashanicus</i>	0.073 ± 0.010
<i>Spermophilus dauricus</i>	2/2	1.00 ± 0.50	0.003 ± 0.002	0.003 ± 0.002	0.003	<i>S. alashanicus</i>	0.074 ± 0.010
<i>Spermophilus erythrognys</i>	4/4	1.00 ± 0.18	0.030 ± 0.009	0.030 ± 0.005	0.044	<i>S. major</i>	0.024 ± 0.004
<i>Spermophilus fulvus</i>	10/5	0.67 ± 0.16	0.003 ± 0.001	0.003 ± 0.001	0.009	<i>S. major</i>	0.032 ± 0.006
<i>Spermophilus major</i>	10/4	0.53 ± 0.18	0.002 ± 0.001	0.002 ± 0.001	0.006	<i>S. brevicauda</i>	0.005 ± 0.002
<i>Spermophilus pallidicauda</i>	2/2	1.00 ± 0.50	0.002 ± 0.001	0.002 ± 0.001	0.002	<i>S. ralli</i>	0.030 ± 0.007
<i>Spermophilus pygmaeus</i>	19/12	0.95 ± 0.03	0.018 ± 0.001	0.019 ± 0.003	0.035	<i>S. pallidicauda</i>	0.071 ± 0.009
<i>Spermophilus ralli</i>	1/1	-	-	-	-	<i>S. major</i>	0.017 ± 0.005
<i>Spermophilus relictus</i>	1/1	-	-	-	-	<i>S. ralli</i>	0.029 ± 0.006
<i>Spermophilus suslicus</i>	15/13	0.97 ± 0.04	0.015 ± 0.004	0.015 ± 0.002	0.040	<i>S. pallidicauda</i>	0.077 ± 0.010
<i>Spermophilus taurensis</i>	1/1	-	-	-	-	<i>S. citellus</i>	0.044 ± 0.008
<i>Spermophilus xanthoprimum</i>	1/1	-	-	-	-	<i>S. taurensis</i>	0.079 ± 0.011
<i>Urocitellus undulatus</i>	13/6	0.77 ± 0.10	0.014 ± 0.004	0.014 ± 0.003	0.035	<i>U. columbianus</i>	0.055 ± 0.008
<i>Urocitellus parryii</i>	2/2	1.00 ± 0.50	0.002 ± 0.001	0.002 ± 0.001	0.002	<i>U. richardsonii</i>	0.024 ± 0.006

n/N – sample size/number of haplotypes; *h* ± SD – haplotype diversity ± standard deviation; π ± SD—nucleotide diversity (per site) ± standard deviation; Mean intra-sp ± SE – mean intraspecific *p*-distance ± standard error; Max intra-sp – maximum intraspecific *p*-distance; Distance to NS ± SE – *p*-distance to a nearest species ± standard error (based on 1000 bootstrap replications).

doi:10.1371/journal.pone.0117201.t001

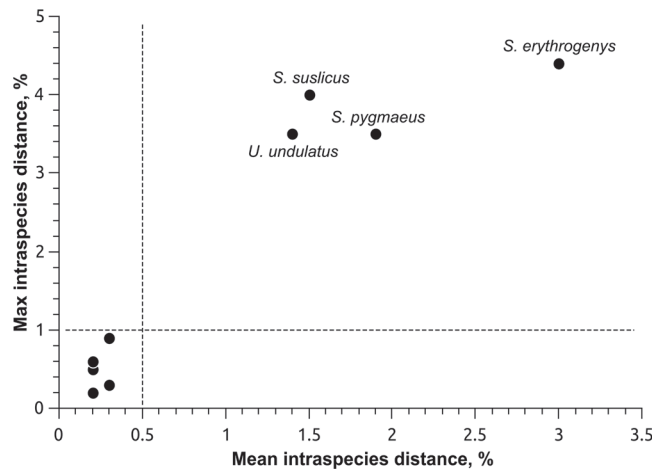


Figure 2. Plot of maximum intraspecies distances against mean intraspecies distances of Eurasian ground squirrels (excluding singletons). Species names are given for putative species complexes indicated by DNA barcoding.

doi:10.1371/journal.pone.0117201.g002

within species with the widest distribution ranges: *S. erythrogegens* (4.4%), *S. suslicus* (4.0%), *S. pygmaeus* (3.5%), and *U. undulatus* (3.5%) (Fig. 2, Table 1).

According to jModelTest 2.1.1 AIC, the best model for our COI dataset was HKY+I+G (Fig. 3). While recognizing the limitations of a single-gene approach and refraining from inferring phylogenetic conclusions, we note that the obtained tree is in agreement with current views on the taxonomy of ground squirrels and corroborate the findings obtained using another mitochondrial marker – cytochrome *b* [11, 25]. The monophyly of *Spermophilus* and *Urocitellus* was supported by high bootstrap values (76–99%; Fig. 3). Within *Spermophilus*, the grouping of species from the subgenus *Colobotis* also had high bootstrap support (98%). This grouping also included *S. relictus* and *S. ralli* which were previously assigned to the subgenus *Citellus* [47] or conf. *Urocitellus* [48]. On the other hand, the monophyly of the subgenus *Citellus* was not supported.

mtDNA introgression in *Spermophilus*

In four ground squirrel species from the Volga region (*S. fulvus*, *S. major*, *S. pygmaeus*, and *S. erythrogegens*) several putative hybridization events were detected (Table 2). In particular, 20% and 13.3% of examined specimens of *S. major* had COI haplotypes of *S. pygmaeus* and *S. fulvus* respectively. In *S. fulvus*, 16.6% of individuals had haplotypes of *S. major*; 5% of *S. pygmaeus* had *S. fulvus* haplotypes; and 20% of *S. erythrogegens* had *S. major* haplotypes.

Table 2. Frequency of the introgressed mtDNA haplotypes in four ground squirrel species as revealed by DNA barcoding.

Species	mtDNA haplotypes		
	<i>S. major</i>	<i>S. fulvus</i>	<i>S. pygmaeus</i>
<i>Spermophilus major</i>	10	2	3
<i>Spermophilus fulvus</i>	2	10	0
<i>Spermophilus pygmaeus</i>	0	1	19
<i>Spermophilus erythrogegens</i>	1	0	0

doi:10.1371/journal.pone.0117201.t002

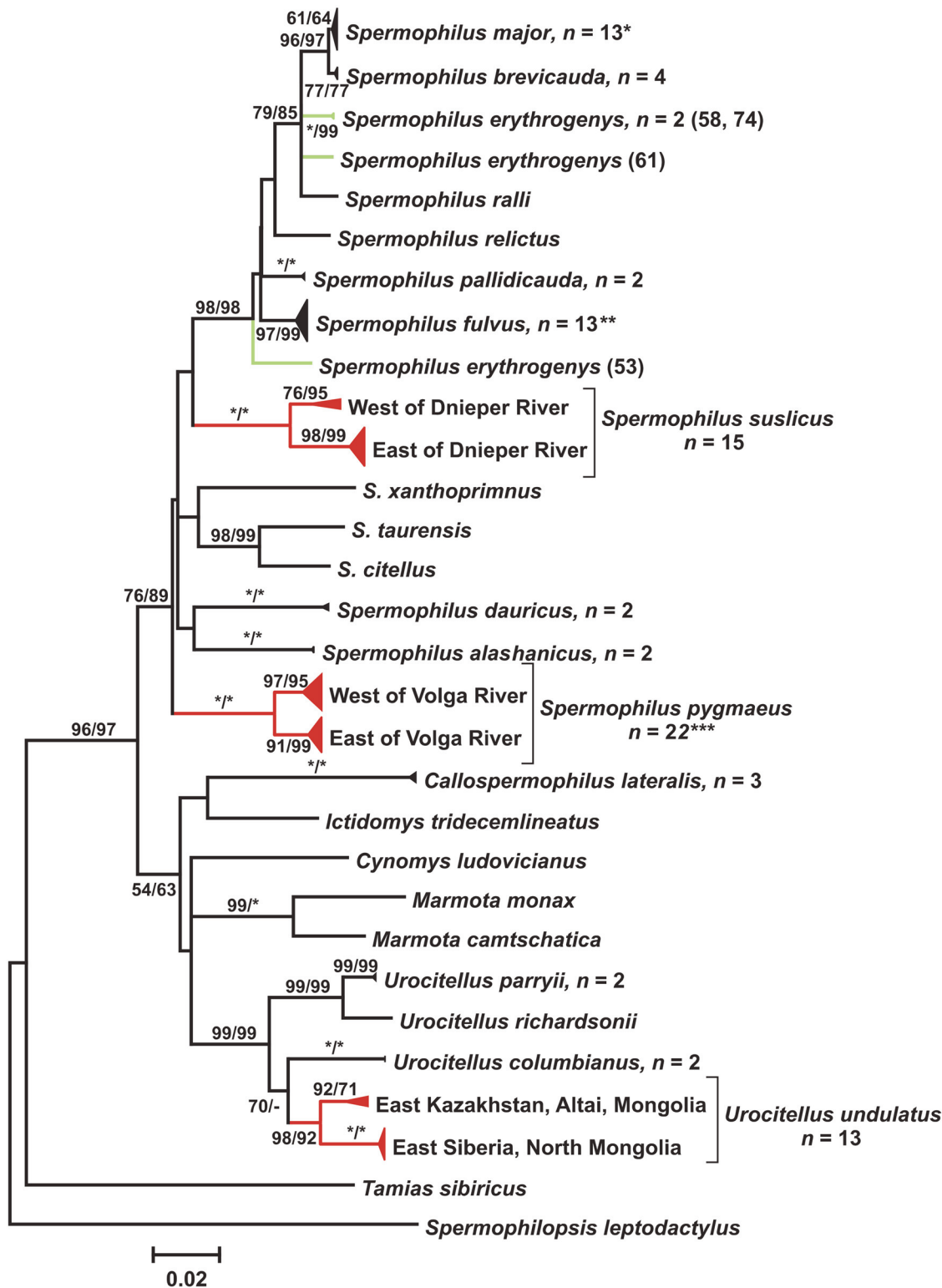


Figure 3. ML tree (HKY+I+G model) for the 657 bp fragment of COI of all Eurasian ground squirrel species. Bootstrap values above 50 are indicated; asterisks represent bootstrap values of 100. The nodes with multiple specimens were collapsed to a triangle, with the horizontal depth indicating the level of intraspecific divergence. Numbers next to each species name indicate the sample size (not indicated when n = 1). * individuals with introgressed mtDNA are not included into this analysis.

doi:10.1371/journal.pone.0117201.g003

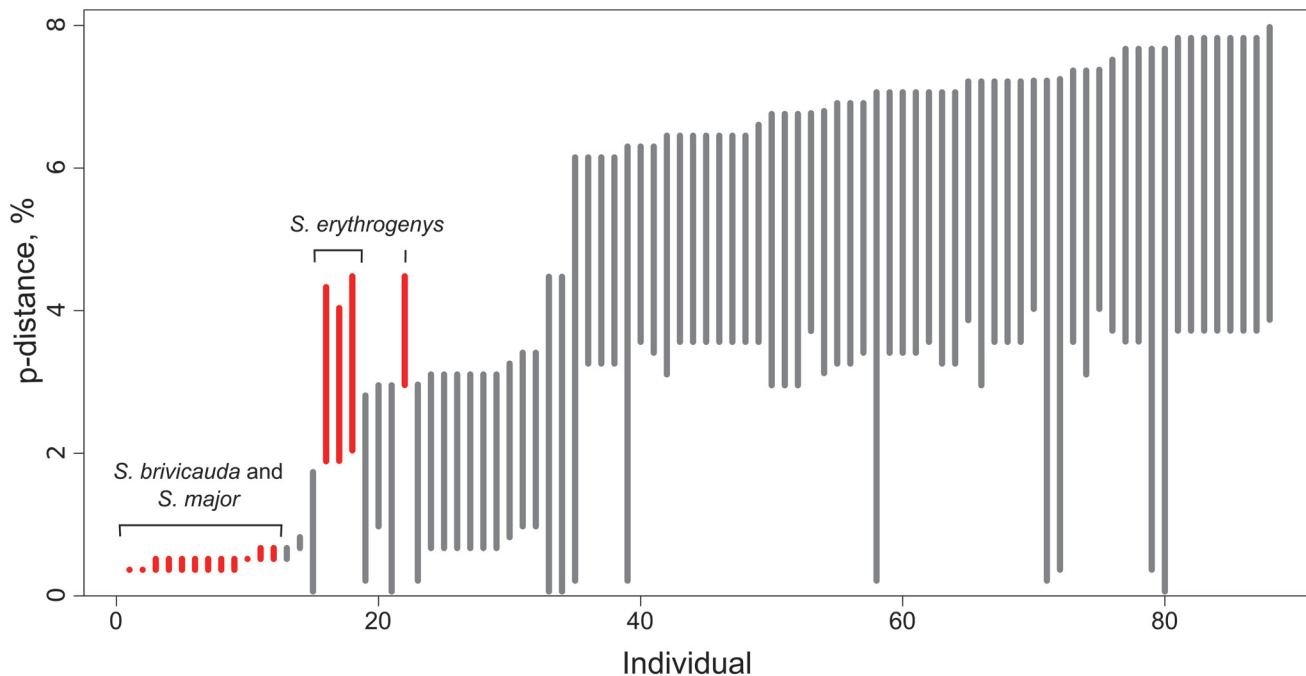


Figure 4. Line-plot of the barcode gap for the Eurasian ground squirrels as generated by Spider [40]. For each individual in the dataset, the grey lines represent the furthest intraspecific distance (bottom of line value), and the closest interspecific distance (top of line value). The red lines show where this relationship is reversed, and the closest non-conspecific is actually closer to the query than its nearest conspecific (i.e. no barcoding gap). Individuals with introgressed mtDNA are not included.

doi:10.1371/journal.pone.0117201.g004

Identification success

The test of barcoding efficiency was performed on the dataset with excluded singletons and individuals harbouring introgressed mtDNA. The “species identification” method (“threshID” in Spider) with a default threshold of 1% allowed the correct identification of 62 individuals, while 14 were ambiguous and seven individuals had no matches (74.7% success rate). The “best close match” approach performed better: it correctly identified 76 individuals (91.6%), while seven ranked as having “no matches”. Optimisation of the threshold value via minimization of false-positive and false-negative errors and using the “localMinima” function showed identical results with an optimal threshold value of 1.7%. Application of the optimal threshold value notably improved identification success with 67 correct, 14 ambiguous and two “no matches” identifications using “threshID” (80.7%) and 81 correct and two “no matches” via “best close match” (97.6%).

The ‘barcode gap’ between species was present in most cases, *i.e.*, maximum intraspecific distances were smaller than minimum interspecific distances for 72 out of 88 individuals in the dataset, including singletons (Fig. 4). There was no ‘barcode gap’ between *S. major* and *S. brivicauda* (due to small interspecies distances) and for *S. erythrogegens* (due to high intraspecies distances).

Automatic species delineation

According to bPTP, the estimated number of species was between 17 and 60, with a mean of 36; however, the best maximum likelihood solution provided a more reliable estimate with 21 species identified (Fig. 5). The RESL algorithm implemented in BOLD identified 20 BIN clusters (Fig. 6), while the number of currently recognized ground squirrel taxa in Eurasia is 16. As

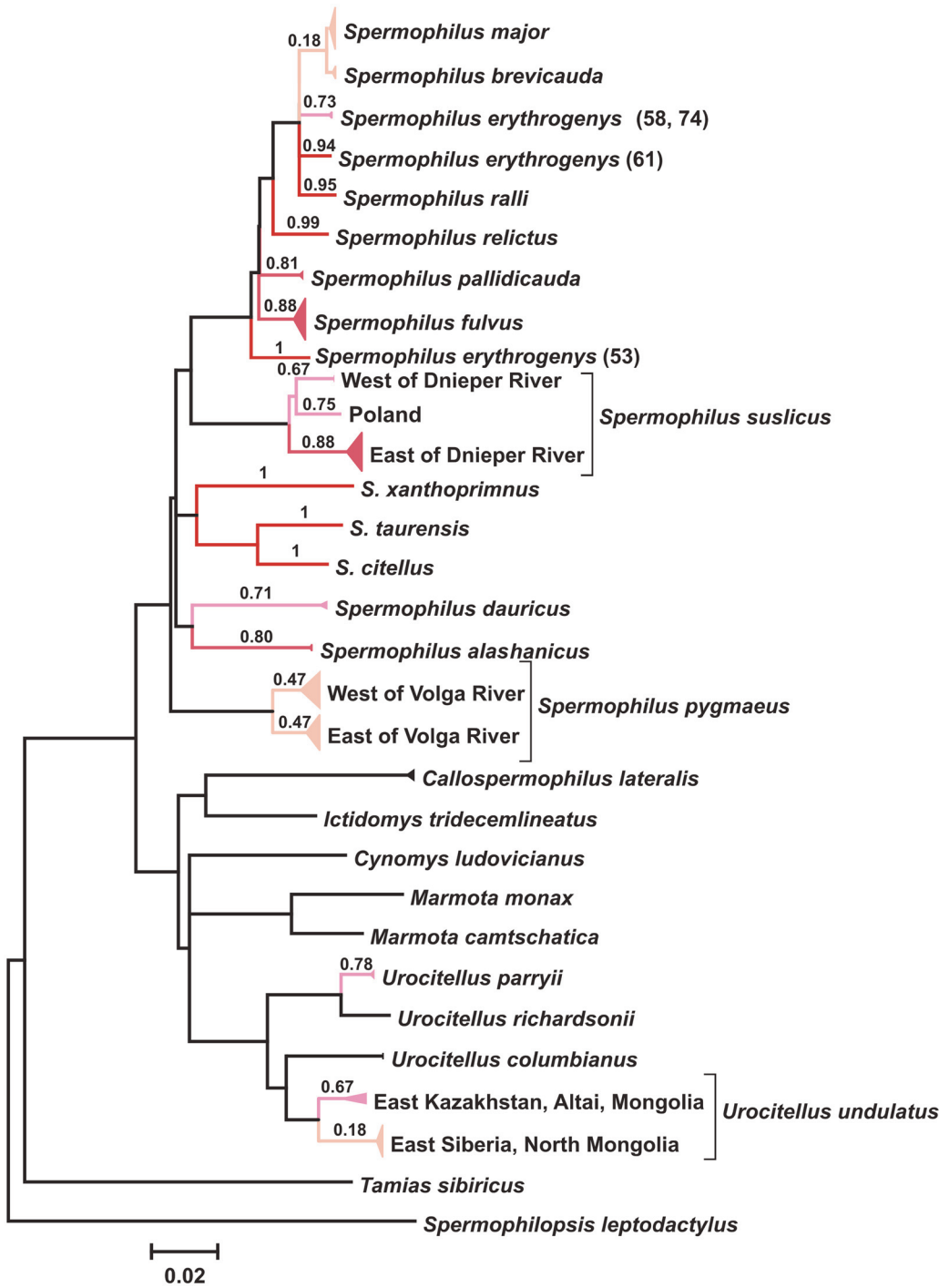


Figure 5. Species delineation of Eurasian ground squirrels by Poisson tree processes (PTP) on the ML tree (HKY+I+G model; 657 bp of COI). Bayesian support values for delimited species are indicated; intensity of red color reflects the strength of support. The nodes with multiple specimens were collapsed to a triangle, with the horizontal depth indicating the level of intraspecific divergence. Individuals with introgressed mtDNA are not included into the tree.

doi:10.1371/journal.pone.0117201.g005

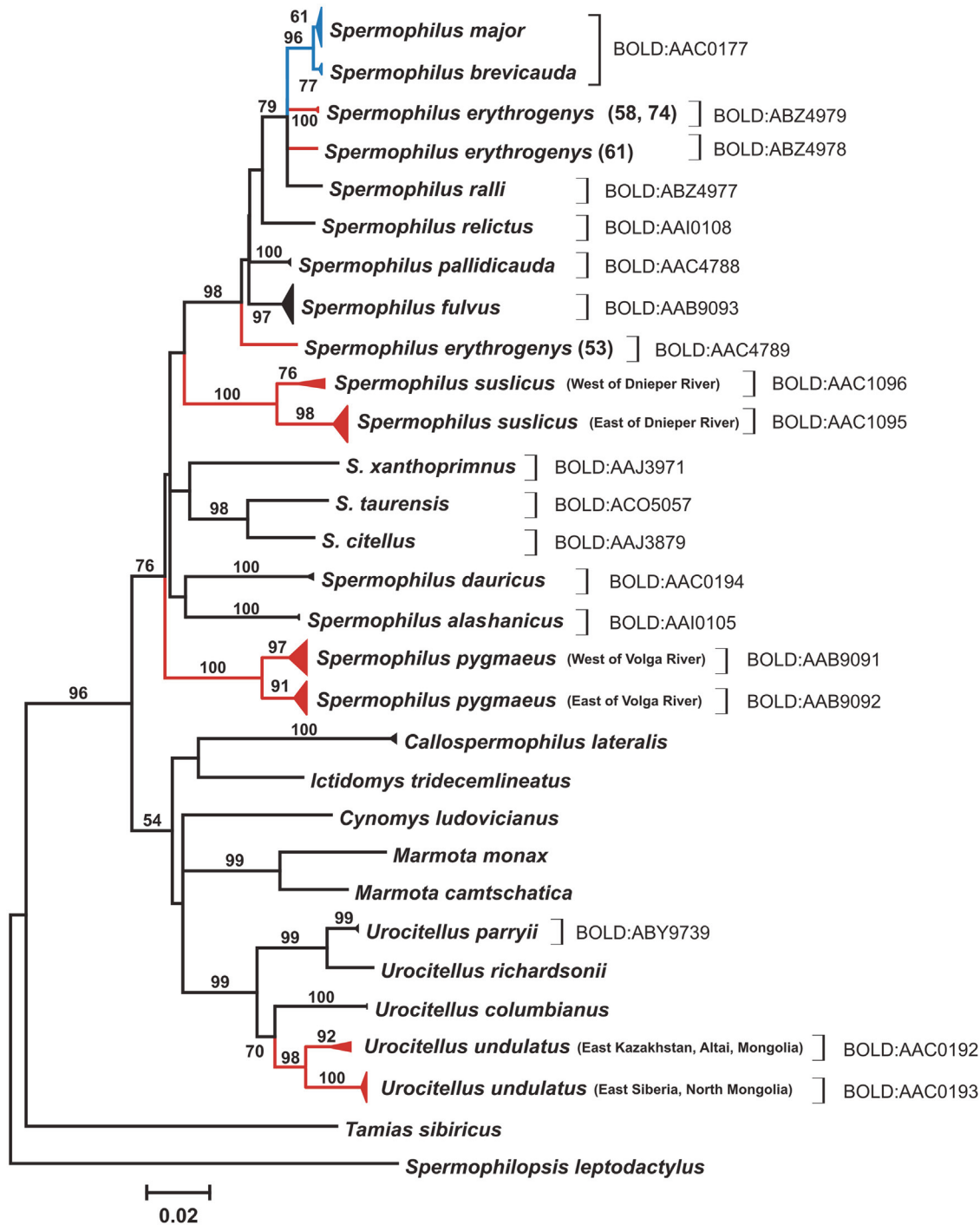


Figure 6. Species delineation of Eurasian ground squirrels by refined single linkage (RESL) analysis showed on the gene tree. Square brackets indicate putative species recognized by RESL. Barcode Index Numbers (BIN) assigned to each putative species are given as BOLD:XXXXXXX. Species split by RESL are red colored, while merged species are blue colored. ML tree (HKY+I+G model) of 657 bp COI fragment. Bootstrap values above 50 are indicated. The nodes with multiple specimens were collapsed to a triangle, with the horizontal depth indicating the level of intraspecific divergence. Individuals with introgressed mtDNA are not included into the tree.

doi:10.1371/journal.pone.0117201.g006

Table 3. Identification success for the best 100 bp and 50 bp mini-barcodes identified by sliding windows analysis in Spider.

Position*	Width, bp	Optimal threshold, %	Best Close Match				Success rate, %
			ambiguous	correct	incorrect	no id	
40	50	2.1	28	52	1	2	62.7
379	50	1.0	13	61	2	7	73.5
46	100	0.9	0	70	0	13	84.3
322	100	1.1	15	66	1	1	79.5

* position of the first nucleotide of the given window in the alignment.

doi:10.1371/journal.pone.0117201.t003

expected, both approaches split species with observed high intraspecific distances (*S. erythro-*
genys, *S. suslicus*, *S. pygmaeus*, and *U. undulatus*) and merged *S. major* and *S. brevicauda*.

Sliding windows analysis

The analysis using 100 bp and 50 bp sliding windows detected two regions within the COI bar-
code sequence, potentially useful for the development of diagnostic ‘mini-barcodes’. Results of
threshold optimisation and estimations of identification success for the four best 100 bp and 50
bp windows are given in Table 3. The best identification success (84.3%) was with the 100 bp
fragment that started from position 46 in the alignment.

Detection of NUMTs

We obtained one putative NUMT (nuclear mitochondrial pseudogene) sequence when analyz-
ing the COI fragment from a fresh ungual phalanx taken from a roadkill *U. undulatus*
(KP098531). The NUMT sequence had an unusually high number of nucleotide substitutions
(8.8%) and codon deletion (positions 274–276). Furthermore, eight out of 13 CpG-dinucleo-
tides (methylation sites) had mutations in the NUMT sequence, while the same dinucleotides
in *U. undulatus* COI sequences carried only three mutations.

Discussion

Species delineation and introgression of mtDNA

COI barcode sequences allowed the correct identification of 12 out of 16 Eurasian ground
squirrel species; the identification of four species (*S. fulvus*, *S. major*, *S. pygmaeus*, and
S. erythro-
genys) was complicated by mtDNA introgression. When individuals with intro-
gressed mtDNA were removed, the best method for species identification was Meier’s “best
close match” criterion that had a 91.6% success rate at a standard threshold of 1%. The proce-
dure of threshold optimization proved to be useful and increased identification success by 6%
in both methods. Notably, we found no differences in the performance of different substitution
models or their advantage over *p*-distance in species delineation.

Interspecific divergence values observed among Eurasian ground squirrels (6.9% in
Spermophilus and 11.3% in *Urocitellus*) are congruent with other COI surveys in vertebrates
[4, 49–52]. The ‘barcode gaps’ were present in most cases, except *S. brevicauda*, *S. major* and
S. erythro-
genys. In the case of *S. brevicauda* and *S. major*, it was caused by shallow interspecific
divergence, while *S. erythro-*
genys turned out to be polyphyletic, comprised of genetically distant
forms (see discussion below). This was reflected in the results of two automatic (OTU-based)
species delineation methods applied to our dataset. Both PTP and RESL analyses merged *S. bre-*
vicauda and *S. major* into single species, while *S. erythro-*
genys was split to three species. The

performance of these species delimitation methods was very similar, distinguishing more putative species (PTP – 21, RESL – 20) than traditionally recognized through morphological taxonomy (16). In addition to *S. erythrogegens*, these analyses also split *S. suslicus*, *S. pygmaeus*, and *U. undulatus*, in concordance with recent findings on morphological, karyotypic and genetic intraspecific variability of these species (see [discussion](#) below).

Barcode-based species recognition was especially hindered in *S. major*, where 33.3% of individuals had COI haplotypes from other species. These findings are not surprising, because the hybridization and mtDNA introgression between these ground squirrel species has been extensively documented [[53–58](#)]. Studies of the variability in the D-loop region within a broader sample has revealed 36.7% (52 out of 137 individuals) of *S. major* to possess haplotypes typical for *S. fulvus* and *S. pygmaeus* [[59](#)]. The same study found no introgressed haplotypes among 119 individuals of the remainder three species, suggesting predominant participation of *S. major* in backcrosses. Thus, in areas of introgression the identification of these species using mtDNA is complicated and the application of COI alone can lead to misclassification.

The small genetic distance (0.5%) and the absence of barcoding gap between two morphologically distinct species – *S. breviceauda* and *S. major* can be explained by ancient hybridization and introgression of mtDNA. We found prominent nuclear – mitochondrial DNA incongruence by analysing three nuclear loci (gene *p53* intron 6, gene *bcr* intron 13, and HoxB gene intron 5 – 1636 bp in total; unpublished data). This analysis suggests that *S. breviceauda* and *S. major* are not sister taxa as could be inferred from mtDNA. The most plausible explanation of this pattern is total replacement of native *S. major* mtDNA in the course of past introgression with *S. breviceauda*, followed by divergence. A similar pattern has been recently described in brown and polar bears [[60](#)], long-tailed and Menzbier's marmots [[61](#)], serotines [[62](#)], and mouse-eared bats, *Myotis* [[63](#)] where ancient hybridization has been inferred from the discordances in mitochondrial and nuclear genome differences.

Introgression events are known to distort the congruence between “species” and “gene” trees [[64](#)] and have been identified as a challenge for DNA barcoding. The inability of COI-based barcodes to correctly distinguish between species due to mtDNA introgression was reported in a number of animal taxa [[10](#), [65](#), [66](#)]; on the other hand, it has been argued that in some of these situations DNA barcodes may still provide adequate resolution for practical identification purposes, provided that the reference library is well populated [[52](#)].

Some researchers proposed to use additional nuclear loci in combination with COI. For instance, Raupach *et al.* [[66](#)] argued that a combination of COI and nuclear ribosomal expansion segments is an efficient tool for identification of carabid (Coleoptera) species. A number of studies utilized the interphotoreceptor retinoid-binding protein for species delimitation of carnivores and rodent species [[8](#), [67–69](#)]. A similar approach could be reliable for delineation of *Spermophilus* species showing mtDNA introgression, since the frequency of “alien” alleles at nuclear loci in these species is considerably lower [[57](#)]. Furthermore, the geographical origin of samples should be taken into account when possible. All *S. major* specimens with introgressed haplotypes originated from an area along the Volga River, approximately 250 km wide. Here, more than 60% *S. major* individuals (n = 96) carry mtDNA from other ground squirrel species [[59](#)].

Intraspecific genetic diversity and divergence

Deep COI barcode divergence was observed within the most widespread ground squirrel species ([Fig. 2](#)) and was highlighted in the results of two OTU-based species delineation approaches used in the study (Figs. [5](#), [6](#)). The range of *S. suslicus* spans 1700 km west to east and is divided by several large rivers [[13](#)]. Three to five subspecies have been previously recognized in this species [[11](#), [47](#), [70](#)]. Furthermore, it is the only species of *Spermophilus* possessing two

chromosome races divided by Dnieper River: western ($2n = 36$, $NF = 72$) and eastern ($2n = 34$, $NF = 68$) [71, 72]. Zagorodnuk and Fedorchenko [73] and Korablev [74] proposed that these two races may constitute two different species. Our analysis has revealed a significant level of divergence (up to 4%) between them, which is concordant with an earlier study of D-loop variability that detected 8% genetic distance between these populations [75]. Notably, the PTP analysis split this taxon into three putative species, distinguishing the westernmost population (from Poland) as a separate species (Fig. 5), while RESL discriminated only two taxa divided by Dnieper River (Fig. 6). These results underscore the need for further taxonomic revision of *S. suslicus* s.l., with possible recognition of *S. suslicus* s. str. in the east and *S. odessanus* Nordman, 1840 in the west as distinct species.

Another widespread species showing the east-west split is *S. pygmaeus*. COI barcodes show genetic divergence of up to 3.5% between populations from the west and east banks of the Volga River. Furthermore, both PTP and RESL analyses distinguished them as putative species (Figs. 5, 6). These findings support our previous observations based on the variability of the D-loop region of mtDNA which revealed a 7% genetic distance between them [56] and suggested the subspecies status of the Caucasian mountain ground squirrel (*S. p. musicus*) [76], previously treated as a distinct species. Deep divergence between these two groups is congruent with paleontological data that shows *S. pygmaeus* to have a substantial level of morphological variability as early as Upper Pleistocene [47].

S. erythrogegens sensu lato also has a wide range and consists of a number of morphologically different forms with unclear systematics and taxonomy. Gromov *et al.* [47] considered *S. erythrogegens* to be a polymorphic species with a number of subspecies, but Ognev [13] and Sludskiy *et al.* [77] distinguished several independent species. COI barcodes show a mean intraspecific *p*-distance of 3.0% (max 4.4%), with splits into several clusters distinguished by both automatic species delimitation approaches used (Figs. 5, 6). These findings support the opinion that *S. erythrogegens* s. l. represents a species complex; however, the relationships between COI haplogroups within it are more complex than the simple east-west splits associated with prominent dispersal barriers. At least three taxonomic units of putative species rank could be identified: (1) *erythrogegens* (right bank of the Irtysh River) (№ 53, Fig. 3); (2) previously unknown form from the right bank of Ob' River (Kuznetsk Depression) (№ 61, Fig. 3); (3) *carruthersi* (Zaysan Depression, Dzungarian Alatau) (№ 58, 74, Fig. 3).

U. undulatus is the most widespread species of the genus *Urocitellus* inhabiting Eurasia. Its ancestors migrated from North America to north Asia in Upper Pliocene – Lower Pleistocene [78]. Six subspecies of *U. undulatus* are currently recognized [11, 47]. Craniometric data [79] and RAPD-PCR [80] suggest that western and eastern subspecies comprise two groupings of species rank: *S. (u.) undulatus* in the east and *S. (u.) eversmanni* in the west, with the border between lying in Lake Baikal area and northern Mongolia [81, 82]. Our DNA barcode data show a 3.5% genetic distance between these groupings, and both PTP and RESL species delineation approaches identified two putative species within this taxon (Figs. 5, 6).

NUMTs

Nuclear copies of mitochondrial DNA (NUMTs) are considered a challenge in using mitochondrial DNA for species diagnosis in DNA barcoding [83–86]. NUMTs can be inadvertently amplified while targeting mitochondrial loci, particularly with broad-range primers, and may bias the final dataset. They have been found in over 64 species [85]. Some have suggested that NUMTs make the barcoding approach unreliable, at least in primates [87]. The low frequency (0.9%) of paralogous nuclear COI sequences in our study suggests that this complication may have been over-stated and does not pose problems with ground squirrels. In our dataset the

observed case of NUMT detection was overcome by re-extraction of DNA followed by amplification and sequencing under usual conditions.

Mini-barcodes

Mini-barcodes are small (ca. 100–400 bp) COI gene fragments that have been proposed for use in cases when full-length barcodes (650 bp) could not be retrieved from archived specimens and processed biological material [88]. Today, the increased use of next-generation sequencing platforms broadened the scope and utility of mini-barcodes for species identification purposes. It has been used for the analysis of soil DNA from past and present ecosystems [89] and for diet assessment (reviewed in [90]). Although many of these studies have used alternative DNA markers (e.g. 12S, 16S, ITS1, *cyt b*), there is a growing body of literature using COI sequencing in NGS approaches (e.g., [5]). Our analysis demonstrated high discrimination ability of 100 bp mini-barcodes for Eurasian ground squirrel species (84.3%) with no incorrect assessments. It underscores the utility of existing COI barcode libraries (www.boldsystems.org) for the development of mini-barcodes and their use in species identification.

Conclusion

Overall, the results of this study provide evidence for the ability of DNA-barcodes to identify most species of Eurasian ground squirrels. Limitations of this approach involve several cases of geographically restricted mtDNA introgression and one case of species polyphyly (in *S. erythrogegens*). The incorporation of nuclear markers and a more fine-grained geographic representation of samples in the reference library should improve identification of *Spermophilus* species engaged in hybridisation and mtDNA introgression. The existence of several genetically divergent haplogroups within several species with wide distribution ranges calls for their in-depth taxonomic reassessment.

Supporting Information

S1 Table. Information on collecting localities, GenBank and BOLD accessions for all specimens used in this study (XLS).
(XLS)

Acknowledgments

Processing specimens at the ZMMU and ZISP was done with administrative support from Igor Pavlinov, Vladimir Lebedev and Fedor Golenishev. Sequencing at the Biodiversity Institute of Ontario, was done with administrative support from Paul Hebert. We thank Vaclav Gvoždík and Oleksandra Godnek for providing additional samples.

Author Contributions

Conceived and designed the experiments: OAE ES SVT AVB. Performed the experiments: OAE ES VLS OVB NVI. Analyzed the data: OAE ES SVT AVB. Contributed reagents/materials/analysis tools: OAE ES VLS SVT OVB NVI AVB. Wrote the paper: OAE ES VLS SVT OVB NVI AVB.

References

1. Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proc R Soc B* 270: 313–321. doi: [10.1098/rspb.2002.2218](https://doi.org/10.1098/rspb.2002.2218) PMID: [12614582](https://pubmed.ncbi.nlm.nih.gov/12614582/)

2. Footitt RG, Maw HEL, von Dohlen CD, Hebert PDN (2008) Species identification of aphids (Insecta: Hemiptera: Aphididae) through DNA barcodes. *Mol Ecol Res* 8: 1189–1201. doi: [10.1111/j.1755-0998.2008.02297.x](https://doi.org/10.1111/j.1755-0998.2008.02297.x)
3. Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, et al. (2008) Identifying Canadian fresh water fishes through DNA barcodes. *PLoS ONE* 3(6): e2490. doi: [10.1371/journal.pone.0002490](https://doi.org/10.1371/journal.pone.0002490) PMID: [22423312](https://pubmed.ncbi.nlm.nih.gov/22423312/)
4. Bitanyi S, Bjørnstad G, Ernest ME, Nesje M, Kusiluka LJ, et al. (2011) Species identification of Tanzanian antelopes using DNA barcoding. *Mol Ecol Res* 11: 442–449. doi: [10.1111/j.1755-0998.2011.02980.x](https://doi.org/10.1111/j.1755-0998.2011.02980.x)
5. Clare EL, Lim BK, Fenton MB, Hebert PDN (2011) Neotropical bats: estimating species diversity with DNA barcodes. *PLoS ONE* 6(7): e22648. doi: [10.1371/journal.pone.0022648](https://doi.org/10.1371/journal.pone.0022648) PMID: [21818359](https://pubmed.ncbi.nlm.nih.gov/21818359/)
6. Rolo EA, Oliveira AR, Dourado CG, Farinha A, Rebelo MT, et al. (2013) Identification of sarcosaprophagous Diptera species through DNA barcoding in wildlife forensics. *Forensic Sci Int* 228(1–3): 160–164. doi: [10.1016/j.forsciint.2013.02.038](https://doi.org/10.1016/j.forsciint.2013.02.038) PMID: [23597753](https://pubmed.ncbi.nlm.nih.gov/23597753/)
7. Borisenko AV, Lim BK, Ivanova NV, Hanner RH, Hebert PDN (2008) DNA barcoding in surveys of small mammal communities: a field study in Suriname. *Mol Ecol Res* 8(3): 471–479. doi: [10.1111/j.1471-8286.2007.01998.x](https://doi.org/10.1111/j.1471-8286.2007.01998.x)
8. Barbosa S, Pauperio J, Searle JB, Alves PC (2013) Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to noninvasive sampling. *Mol Ecol Res* 13(1): 43–56. doi: [10.1111/1755-0998.12024](https://doi.org/10.1111/1755-0998.12024)
9. Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012) Environmental DNA. *Mol Ecol* 21: 1789–1793. doi: [10.1111/j.1365-294X.2012.05542.x](https://doi.org/10.1111/j.1365-294X.2012.05542.x) PMID: [22486819](https://pubmed.ncbi.nlm.nih.gov/22486819/)
10. Nesi N, Nakouné E, Cruaud C, Hassanin A (2011) DNA barcoding of African fruit bats (Mammalia, Pteropodidae). The mitochondrial genome does not provide a reliable discrimination between *Epomophorus gambianus* and *Micropteropus pusillus*. *C R Biol* 334(7): 544–554. doi: [10.1016/j.crvi.2011.05.003](https://doi.org/10.1016/j.crvi.2011.05.003) PMID: [21784364](https://pubmed.ncbi.nlm.nih.gov/21784364/)
11. Helgen KM, Cole FR, Helgen LE, Wilson DE (2009) Generic Revision in the Holarctic Ground Squirrel Genus *Spermophilus*. *J Mammal* 90: 270–305. doi: [10.1644/07-MAMM-A-309.1](https://doi.org/10.1644/07-MAMM-A-309.1)
12. Howell AH (1938) Revision of the North American ground squirrels with a classification of the North American Sciuridae. *North Am Fauna* 56: 1–256. doi: [10.3996/nafa.56.0001](https://doi.org/10.3996/nafa.56.0001)
13. Ognev SI (1947) Mammals of the U.S.S.R. and adjacent countries. Vol. V. Rodents. Published for the Smithsonian Institution and the National Science Foundation, Washington, D.C., by the Israel Program for Scientific Translations, Jerusalem (English translation published 1963).
14. Afonin AN, Greene SL, Dzyubenko NI, Frolov AN (eds.) (2008) Interactive Agricultural Ecological Atlas of Russia and Neighboring Countries. Economic Plants and their Diseases, Pests and Weeds. Available at: <http://www.agroatlas.ru>. Accessed 11 August 2014.
15. Kucheruk VV (1977) Wild mammals as carriers of diseases dangerous to man. In: *Advances in modern theriology*. Moscow: Nauka. pp. 75–92. In Russian.
16. Verzhutskii DB (1999) The epizootiological role of the population organization of the stock of fleas on the long-tailed suslik in a natural focus of plague in Tuva. *Parazitologiya* 33(3): 242–250. In Russian. PMID: [10771772](https://pubmed.ncbi.nlm.nih.gov/10771772/)
17. Bazanova LP, Innokent'eva TI (2012) Reservation forms of plague infectious agent in Tuva natural focus. *Zhurnal Mikrobiol Epidemiol Immunobiol Sep-Oct* (5): 115–119. In Russian.
18. Shilova SA, Tchabovsky AV (2009) Population response of rodents to control with rodenticides. *Curr Zool* 55(2): 81–91.
19. Koshev YS (2008) Distribution and status of the European Ground Squirrel (*Spermophilus citellus*) in Bulgaria. *Lynx (Praha)* 39(2): 251–261.
20. Shekarova ON, Neronov VV, Savinetskaya LE (2008) Speckled ground squirrel (*Spermophilus suslicus*): current distribution, population dynamics and conservation. *Lynx (Praha)* 39: 317–322.
21. Enzinger K, Walder C, Moser D, Holzer T, Zulka P, et al. (2012) The ground-squirrel action plan of Lower Austria: results from eight years of suslik conservation in the Austrian province. In: *Abstr IV Eur Ground Squirrel Meeting*. Poznan: Pol Soc Nat Conserv “Salamandra”. pp. 19.
22. Tokaji K, Váczi O, Bakó B, Gedeon CI (2012) 25 years of translocation programmes on EGS in Hungary. In: *Abstr IV European Ground Squirrel Meeting*. Poznan: Pol Soc Nat Conserv “Salamandra”. pp. 17.
23. The IUCN Red List of Threatened Species. Version 2014. 2. Available at: <http://www.iucnredlist.org>. Accessed 2014 Aug 11.

24. Coroiu C, Kryštufek B, Vohralík V, Zagorodnyuk I (2008) *Spermophilus citellus*. The IUCN Red List of Threatened Species. Version 2014.2. Available at: <http://www.iucnredlist.org/details/20472/0>. Accessed 11 August 2014.
25. Harrison RG, Bogdanowicz SM, Hoffmann RS, Yensen E, Sherman PW (2003) Phylogeny and evolutionary history of Ground Squirrels (Rodentia: Marmotinae). *J Mamm Evol* 10(3): 249–276. doi: [10.1023/B:JOMM.0000015105.96065.f0](https://doi.org/10.1023/B:JOMM.0000015105.96065.f0)
26. Herron MD, Castoe TA, Parkinson CL (2004) Sciurid phylogeny and the paraphyly of Holarctic ground squirrels (*Spermophilus*). *Mol Phylogenet Evol* 31: 1015–1030. doi: [10.1016/j.ympev.2003.09.015](https://doi.org/10.1016/j.ympev.2003.09.015) PMID: [15120398](https://pubmed.ncbi.nlm.nih.gov/15120398/)
27. Gunduz I, Jaarola M, Tez C, Yeniurt C, Polly D, et al. (2007) Multigenic and morphometric differentiation of ground squirrels (*Spermophilus*, Scuriidae, Rodentia) in Turkey, with a description of a new species. *Mol Phylogenet Evol* 43: 916–935. doi: [10.1016/j.ympev.2007.02.021](https://doi.org/10.1016/j.ympev.2007.02.021) PMID: [17500011](https://pubmed.ncbi.nlm.nih.gov/17500011/)
28. Ozkurt SO, Sozen M, Yigit N, Kandemir I, Colak R, et al. (2007) Taxonomic status of the genus *Spermophilus* (Mammalia: Rodentia) in Turkey and Iran with description of a new species. *Zootaxa* 1529: 1–15.
29. Ivanova NV, deWaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Mol Ecol Notes* 6: 998–1002. doi: [10.1111/j.1471-8286.2006.01428.x](https://doi.org/10.1111/j.1471-8286.2006.01428.x)
30. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual*. New York, Cold Spring Harbor Laboratory Press. 1626 p.
31. Hall TA (1999) BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41: 95–98.
32. Helaers R, Milinkovitch MC (2010) MetaPIGA v2.0: maximum likelihood large phylogeny estimation using the metapopulation genetic algorithm and other stochastic heuristics. *BMC Bioinform* 11: 379. doi: [10.1186/1471-2105-11-379](https://doi.org/10.1186/1471-2105-11-379)
33. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16: 111–120. doi: [10.1007/BF01731581](https://doi.org/10.1007/BF01731581) PMID: [7463489](https://pubmed.ncbi.nlm.nih.gov/7463489/)
34. Collins RA, Boykin LM, Cruickshank RH, Armstrong KF (2012) Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification. *Methods Ecol Evol* 3: 457–465. doi: [10.1111/j.2041-210X.2011.00176.x](https://doi.org/10.1111/j.2041-210X.2011.00176.x)
35. Srivathsan A, Meier R (2012) On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. *Cladistics* 28: 190–194. doi: [10.1111/j.1096-0031.2011.00370.x](https://doi.org/10.1111/j.1096-0031.2011.00370.x)
36. Posada D (2009) Selection of models of DNA evolution with jMODELTEST. *Bioinform DNA Seq Analysis* 537: 93–112. doi: [10.1007/978-1-59745-251-9_5](https://doi.org/10.1007/978-1-59745-251-9_5)
37. Rambaut A (2008) FigTree v1.4: Tree figure drawing tool. Available: <http://tree.bio.ed.ac.uk/software/figtree>. Accessed 2013 Feb 20.
38. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28: 2731–2739. doi: [10.1093/molbev/msr121](https://doi.org/10.1093/molbev/msr121) PMID: [21546353](https://pubmed.ncbi.nlm.nih.gov/21546353/)
39. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinform* 25: 1451–1452. doi: [10.1093/bioinformatics/btp187](https://doi.org/10.1093/bioinformatics/btp187)
40. Brown SDJ, Collins RA, Boyer S, Lefort M-C, Malumbres-Olarte J, et al. (2012) Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Mol Ecol Res* 12(3): 562–565. doi: [10.1111/j.1755-0998.2011.03108.x](https://doi.org/10.1111/j.1755-0998.2011.03108.x)
41. R Development Core Team (2012) *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available: <http://www.R-project.org>. Accessed 2014 Feb 13.
42. Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Syst Biol* 55: 715–728. doi: [10.1080/10635150600969864](https://doi.org/10.1080/10635150600969864) PMID: [17060194](https://pubmed.ncbi.nlm.nih.gov/17060194/)
43. Meyer CP, Paulay G (2005) DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biol* 3(12): e422. doi: [10.1371/journal.pbio.0030422](https://doi.org/10.1371/journal.pbio.0030422) PMID: [16336051](https://pubmed.ncbi.nlm.nih.gov/16336051/)
44. Meier R, Egert U, Aertsen A, Nawrot MP (2008) FIND – a unified framework for neural data analysis. *Neural Network* 21(8): 1085–1093. doi: [10.1016/j.neunet.2008.06.019](https://doi.org/10.1016/j.neunet.2008.06.019)
45. Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29(22): 2869–2876. doi: [10.1093/bioinformatics/btt499](https://doi.org/10.1093/bioinformatics/btt499) PMID: [23990417](https://pubmed.ncbi.nlm.nih.gov/23990417/)

46. Ratnasingham S, Hebert PDN (2013) A DNA-based registry for all animal species: the barcode index number (BIN) system. PLoS ONE 8(7): e66213. doi: [10.1371/journal.pone.0066213](https://doi.org/10.1371/journal.pone.0066213) PMID: [23861743](https://pubmed.ncbi.nlm.nih.gov/23861743/)
47. Gromov IG, Bibikov DI, Kalabuchov NI, Meier MN (1965) Ground Squirrels (Marmotinae). Fauna USSR, Mammals. V. 3. Iss. 2. Moscow: Nauka. 467 p. In Russian.
48. Gromov IM, Erbaeva MA (1995) Mammals of the fauna of Russia and adjacent regions. Lagomorphs and rodents. (Guides on the Russian fauna published by Zoological Institute of RAS, 167). Saint Petersburg: Zool Inst RAS. 529 p. In Russian.
49. Clare EL, Lim BK, Engstrom MD, Eger JL, Hebert PDN (2007) DNA barcoding of Neotropical bats: species identification and discovery within Guyana. Mol Ecol Notes 7(2): 184–190. doi: [10.1111/j.1471-8286.2006.01657.x](https://doi.org/10.1111/j.1471-8286.2006.01657.x)
50. Lakra WS, Goswami M, Gopalakrishnan A (2009) Molecular identification and phylogenetic relationships of seven Indian Sciaenids (Pisces: Perciformes, Sciaenidae) based on 16S rRNA and cytochrome oxidase subunit I mitochondrial genes. Mol Biol Rep 36: 831–839. doi: [10.1007/s11033-008-9252-1](https://doi.org/10.1007/s11033-008-9252-1) PMID: [18415704](https://pubmed.ncbi.nlm.nih.gov/18415704/)
51. Eaton M, Meyers G, Kolokotronis S-O, Leslie M, Martin A, et al. (2010) Barcoding bushmeat: molecular identification of Central African and South American harvested vertebrates. Conserv Genet 11: 1389–1404. doi: [10.1007/S10592-009-9967-0](https://doi.org/10.1007/S10592-009-9967-0)
52. Kruskop SV, Borisenko AV, Ivanova NV, Lim BK, Eger JL (2012) Genetic diversity of northeastern palaeartic bats as revealed by DNA barcodes. Acta Chiropt 14: 1–14. doi: [10.3161/150811012X654222](https://doi.org/10.3161/150811012X654222)
53. Bazhanov VS (1944) Ground squirrel hybrids (to the problem of interspecific hybridization in nature). Dokl Biol Sci 12: 321–322. In Russian.
54. Denisov VP (1963) On hybridization of species of the genus Citellus Oken. Zool Zhurnal 42: 1887–1889. In Russian.
55. Nikolsky AA, Starikov VP (1997) Variability of alarm call in *Spermophilus major* and *Spermophilus erythrognys* (Rodentia, Sciuridae) within contact zone in Kurgan district. Zool Zhurnal 76: 845–857. In Russian.
56. Ermakov OA, Surin VL, Titov SV, Tagiev AF, Luk'yanenko AV, et al. (2002) A molecular genetic study of hybridization in four species of Ground Squirrels (*Spermophilus*: Rodentia, Sciuridae). Russ J Genet 38(7): 996–809. doi: [10.1023/A:1016395722664](https://doi.org/10.1023/A:1016395722664)
57. Ermakov OA, Surin VL, Titov SV, Zborovsky SS, Formozov NA (2006) A search for Y-chromosomal species-specific markers and their use for hybridization analysis in Ground Squirrels (*Spermophilus*: Rodentia, Sciuridae). Russ J Genet 42(4): 429–438. doi: [10.1134/S1022795406040107](https://doi.org/10.1134/S1022795406040107)
58. Spiridonova LN, Chelomina GN, Starikov VP, Korablev VP, Zvirka MV, et al. (2005) RAPD-PCR analysis of ground squirrels from the Tobol–Ishim Interfluvium: evidence for interspecific hybridization between ground squirrel species *Spermophilus major* and *S. erythrognys*. Russ J Genet 41(9): 991–1001. doi: [10.1007/s11177-005-0190-X](https://doi.org/10.1007/s11177-005-0190-X)
59. Ermakov OA, Titov SV, Surin VL, Formozov NA (2006) Molecular genetic study of maternal and paternal lineages of hybridization of Ground Squirrels (*Spermophilus*: Rodentia, Sciuridae). Bull Moscow Soc Naturalist 111: 30–35. In Russian.
60. Hailer FV, Kutschera E, Hallstrom BM, Klassert D, Fain SR, et al. (2012) Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. Science 336: 344–347. doi: [10.1126/science.1216424](https://doi.org/10.1126/science.1216424) PMID: [22517859](https://pubmed.ncbi.nlm.nih.gov/22517859/)
61. Brandler OV, Lyapunova EA, Bannikova AA, Kramerov DA (2010) Phylogeny and systematics of marmots (*Marmota*, Sciuridae, Rodentia) inferred from inter-SINE PCR data. Russ J Genet 46: 283–292. doi: [10.1134/S102279541003004X](https://doi.org/10.1134/S102279541003004X)
62. Artyushin IV, Bannikova AA, Lebedev VS, Kruskop SV (2009) Mitochondrial DNA relationships among North Palaeartic Eptesicus (*Vesperilionidae*, Chiroptera) and past hybridization between Common Serotine and Northern Bat. Zootaxa 2262: 40–52.
63. Furman A, Çoraman E, Çelik YE, Postawa T, Bachanek J, et al. (2014) Cytonuclear discordance and the species status of *Myotis myotis* and *Myotis blythii* (Chiroptera). Zool Scr 43: 549–561. doi: [10.1111/zsc.12076](https://doi.org/10.1111/zsc.12076)
64. Petit RJ, Excoffier L (2009) Gene flow and species delimitation. Trends Ecol Evol 24: 386–393. doi: [10.1016/j.tree.2009.02.011](https://doi.org/10.1016/j.tree.2009.02.011) PMID: [19409650](https://pubmed.ncbi.nlm.nih.gov/19409650/)
65. Paquin P, Hedin MC (2004) The power and perils of “molecular taxonomy”: a case study of eyeless and endangered Cicurina (Araneae: Dictynidae) from Texas caves. Mol Ecol 13: 3239–3255. doi: [10.1111/j.1365-294X.2004.02296.x](https://doi.org/10.1111/j.1365-294X.2004.02296.x) PMID: [15367136](https://pubmed.ncbi.nlm.nih.gov/15367136/)

66. Raupach MJ, Astrin JJ, Hannig K, Peters MK, Stoeckle MY, et al. (2010) Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Front Zool* 7: 26. doi: [10.1186/1742-9994-7-26](https://doi.org/10.1186/1742-9994-7-26) PMID: [20836845](https://pubmed.ncbi.nlm.nih.gov/20836845/)
67. Oliveira R, Castro D, Godinho R, Luikart G, Alves PC (2010) Species identification using a small nuclear gene fragment: application to sympatric wild carnivores from South-western Europe. *Conserv Genet* 11: 1023–1032. doi: [10.1007/s10592-009-9947-4](https://doi.org/10.1007/s10592-009-9947-4)
68. Chaval Y, Dobigny G, Michaux J, Pagès M, Corbisier C, et al. (2010) A multi-approach survey as the most reliable tool to accurately assess biodiversity: an example of thai murine rodents. *Kasetsart J Nat Sci* 44: 590–603.
69. Pages M, Chaval Y, Herbreteau V, Waengsothorn S, Cosson JF et al. (2010) Revisiting the taxonomy of the Rattini tribe: a phylogeny-based delimitation of species boundaries. *BMC Evol Biol* 10: 184. doi: [10.1186/1471-2148-10-184](https://doi.org/10.1186/1471-2148-10-184) PMID: [20565819](https://pubmed.ncbi.nlm.nih.gov/20565819/)
70. Thorington RW, Hoffmann RS (2005) Family Sciuridae / Mammal species of the world: a taxonomic and geographic reference in (Wilson D.E. and Reeder D.M., eds.). 3rd ed. Johns Hopkins University Press, Baltimore, Maryland. pp. 754–818.
71. Denisov V, Bielianin A, Jordan M, Rudek Z (1969) Karyological investigations of two species *Citellus* (*Citellus pygmaeus* Pall. and *Citellus suslicus* Guld.). *Folia Biol* 17: 169–174.
72. Liapunova EA, Vorontsov NN (1970) Chromosomes and some issues of the evolution of the ground squirrel genus *Citellus* (Rodentia: Sciuridae). *Experientia* 26: 1033–1038. doi: [10.1007/BF02114181](https://doi.org/10.1007/BF02114181)
73. Zagorodnuk IV, Fedorchenko OO (1995) Allopatric species among rodents of the group *Spermophilus suslicus* (Mammalia). *Vestn Zool* 29(5/6): 49–58. In Russian.
74. Korablev VP (1997) Distribution of chromosomal forms of the speckled ground squirrel *Spermophilus suslicus* Guld., 1770. In: Rare mammal species of Russia and adjacent regions. Abstr Intern Conf, Apr 9–11, 1997, Moscow. Moscow: A.N. Severtzov Inst Ecol Evol RAS. pp. 50. In Russian.
75. Ermakov OA, Surin VL, Titov SV (2011) Genetic diversity and differentiation of the Speckled Ground Squirrel inferred from sequencing of mtDNA control region. *Izv Penz Gos Pedagog Univ im V.G. Belinskogo* 25: 176–180. In Russian.
76. Ermakov OA, Titov SV, Savinetsky AB, Surin VL, Zborovsky SS, et al. (2006) Molecular genetic and palaeoecological arguments for conspecificity of Little (*Spermophilus pygmaeus*) and Caucasian Mountain (*S. musicus*) Ground Squirrels. *Zool Zhurnal* 85: 1474–1483. In Russian.
77. Sludskiy AA, Varshavsky SN, Ismagilov MI, Kapitonov VI, Shubin IG (1969) Mammals of Kazakhstan. Rodents (Marmots and Ground squirrels). Alma-Ata: Nauka of the Kazah SSR. 455p. In Russian.
78. Agadzhanyan AK (2006) Stages of evolution of Ground Squirrels in Northern Eurasia. *Bull Moscow Soc Naturalist* 111: 4–17. In Russian.
79. Linetskaya ON, Linetskiy AI (1989) Species differences in the character of craniometric variability in ground squirrels Eastern Palaearctic. In: Modern approaches to the study of variability. Vladivostok: Far Easter Branch RAS. pp. 99–105. In Russian.
80. Tsvirka MV, Korablev VP (2012) Genetic variability and differentiation of long-tailed ground squirrel (*Spermophilus undulatus*) based on RAPD-PCR analysis. *Tomsk State Univ J Biol* 4: 145–161. In Russian.
81. Vorontsov NN, Frisman LV, Lyapunova EA, Mezheva ON, Serdyk VA, et al. (1980) The effect of isolation on the morphological and genetical divergence of population. *Genetica (Hague)* 52/53: 229–238.
82. Kapustina SY, Brandler OV, Ad'yaa Y (2011) Molecular genetic differentiation of long-tailed ground squirrel (*Urocitellus undulatus*, Marmotinae, Sciuridae). In: Theriofauna of Russia and adjacent territories: Intern Symp. Moscow: KMK. pp. 197. In Russian.
83. van der Kuyl AC, Kuiken CL, Dekker JT, Perizonius WR, Goudsmit J (1995) Nuclear counterparts of the cytoplasmic mitochondrial 12S rRNA gene: a problem of ancient DNA and molecular phylogenies. *J Mol Evol* 40: 652–657. doi: [10.1007/BF00160513](https://doi.org/10.1007/BF00160513) PMID: [7543951](https://pubmed.ncbi.nlm.nih.gov/7543951/)
84. Sorenson MD, Quinn TW (1998) Numts: a challenge for avian systematics and population biology. *Auk* 115: 214–221. doi: [10.2307/4089130](https://doi.org/10.2307/4089130)
85. Bensasson D, Zhang DX, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol Evol* 16: 314–321. doi: [10.1016/S0169-5347\(01\)02151-6](https://doi.org/10.1016/S0169-5347(01)02151-6) PMID: [11369110](https://pubmed.ncbi.nlm.nih.gov/11369110/)
86. Hazkani-Covo E, Zeller RM, Martin W (2010) Molecular Poltergeists: Mitochondrial DNA Copies (numts) in Sequenced Nuclear Genomes. *PLoS Genet* 6: e1000834. doi: [10.1371/journal.pgen.1000834](https://doi.org/10.1371/journal.pgen.1000834) PMID: [20168995](https://pubmed.ncbi.nlm.nih.gov/20168995/)
87. Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc Natl Acad Sci USA* 105: 13486–13491. doi: [10.1073/pnas.0803076105](https://doi.org/10.1073/pnas.0803076105) PMID: [18757756](https://pubmed.ncbi.nlm.nih.gov/18757756/)

88. Meusnier I, Singer GAC, Landry J-F, Hickey DA, Hebert PDN, et al. (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* 9: 214. doi: [10.1186/1471-2164-9-214](https://doi.org/10.1186/1471-2164-9-214) PMID: [18474098](https://pubmed.ncbi.nlm.nih.gov/18474098/)
89. Epp LS, Boessenkool S, Bellemain EP, Haile J, Esposito A, et al. (2012) New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Mol Ecol* 21(8): 1821–1833. doi: [10.1111/j.1365-294X.2012.05537.x](https://doi.org/10.1111/j.1365-294X.2012.05537.x) PMID: [22486821](https://pubmed.ncbi.nlm.nih.gov/22486821/)
90. Pompanon F, Deagle BE, Symondson WOC, Brown DS, Jarman SN, et al. (2012) Who is eating what: diet assessment using next generation sequencing. *Mol Ecol* 21(8): 1931–1950. doi: [10.1111/j.1365-294X.2011.05403.x](https://doi.org/10.1111/j.1365-294X.2011.05403.x) PMID: [22171763](https://pubmed.ncbi.nlm.nih.gov/22171763/)