

Phylogeographical study reveals high genetic diversity in a widespread desert rodent, *Dipus sagitta* (Dipodidae: Rodentia)

VLADIMIR S. LEBEDEV^{1*}, ANNA A. BANNIKOVA², LIANG LU³, EVGENY A. SNYTIKOV², YANSANJAV ADIYA⁴, EVGENIYA N. SOLOVYEVA¹, ALEXEI V. ABRAMOV⁵, ALEXEI V. SUROV⁶ and GEORGY I. SHENBROT⁷

¹Zoological Museum, Moscow State University, B.Nikitskaya 6, 125009 Moscow, Russia

²Lomonosov Moscow State University, Vorobievsky Gory, 119991 Moscow, Russia

³State Key Laboratory for Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Centre of Disease Control and Prevention, Beijing 102206, China

⁴Institute of General and Experimental Biology of Mongolian Academy of Science, Ulaanbaatar 13330, Mongolia

⁵Zoological Institute, Russian Academy of Sciences, Universitetskaya nab. 1, 199034 St. Petersburg, Russia

⁶Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Leninskii pr., 33, Moscow 119071, Russia

⁷Mitrani Department of Desert Ecology, Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion 84990, Israel

Received 30 April 2017; revised 13 July 2017; accepted for publication 14 July 2017

A phylogeographical study of the northern three-toed jerboa, *Dipus sagitta*, which has one of the largest geographical ranges among Palaearctic desert rodents, was performed using complete mitochondrial cytochrome *b* (*Cytb*) and fragments of two nuclear genes. Phylogenetic analysis of the *Cytb* data, including 222 specimens from 69 localities revealed, six allopatric lineages divergent at 7.4–10.1%. In some lineages the *Cytb* data are supported by nuclear data, thus confirming a high level of cryptic diversity within *D. sagitta*. The recovered pattern of genetic differentiation was reviewed from a taxonomic perspective. In general, the mitochondrial lineages show relatively good correlation with morphological subspecies. Both mitochondrial and nuclear data support the isolated position of the Qaidam lineage, which diverged from the rest in the Early Pleistocene (~1.5 Myr) according to our molecular clock results. Therefore, we propose to recognize this lineage as a separate species, *Dipus deasyi*; the taxonomic status of other lineages remains to be clarified. The results are consistent with the hypothesis that the centre of the origin of the *D. sagitta* complex is located in the eastern part of the range. The geographical structure of genetic variation in *D. sagitta* is compared to those in other desert rodents.

ADDITIONAL KEYWORDS: Central Asian deserts – Dipodidae – molecular dating – phylogeography – species delimitation.

INTRODUCTION

Recent molecular studies have demonstrated that species of small mammals with large geographical ranges

often are the complexes of cryptic species (Neronov *et al.*, 2009; Ben Faleh *et al.*, 2012b; Paupério *et al.*, 2012; Petrova *et al.*, 2016) or at least the complexes of deeply divergent phylogenetic lineages within one species (Jaarola & Searl, 2002; Heckel *et al.*, 2005; Ben Faleh *et al.*, 2012a). Usually, geographical structuring of genetic variation across the range of a species can

*Corresponding author. E-mail: wslebedev@hotmail.com

be interpreted as an echo of its past changes in distribution described in terms of extinction across large areas, survival in small refugia and subsequent re-colonization. Such cases are quite common among small mammals of the forest and steppe zones of Eurasia (Michaux, Libois & Filippucci, 2005; McDevitt *et al.*, 2010; Petrova *et al.*, 2015). With regard to Palaearctic desert rodents, studies of genetic variation across geographical ranges of widely distributed species are relatively uncommon. The Great Palaearctic Desert belt is subdivided into several semi-isolated areas (Shenbrot, Krasnov & Rogovin, 1999), and thus cryptic speciation can be expected at the boundaries between such geographical subdivisions. Examples (*Jaculus jaculus s.l.* in the Sahara and Arabia – Ben Faleh *et al.*, 2012b; *Meriones meridianus s.l.* in Kazakhstan, Turan and Central Asia – Neronov *et al.*, 2009) confirmed this expectation. On the other hand, among Palaearctic desert rodents, genetic variation within large areas not subdivided by barriers is generally not structured geographically (Ben Faleh *et al.*, 2012b; Lv *et al.*, 2016; Ndiaye *et al.*, 2016).

The northern three-toed jerboa, *Dipus sagitta* Pallas, 1773, has one of the largest geographical ranges among Palaearctic desert rodents (Fig. 1). It occurs from the Don River valley in the southern part of European Russia in the west (43°E) to the Nen Jiang

River and Song Hua River valleys in north-eastern China in the east (124°E) and from strip pine forests of Altai in the north (52°N) to the deserts of northern Iran, southern Turkmenistan, and the Tarim and Qaidam Basins in the south (36°N). Across this extensive geographical area, the species is well differentiated morphologically into 15–17 subspecies that can be combined into two groups: western (south of European Russia, Kazakhstan and Turan) and eastern (Central Asia) (Shenbrot, 1991a, b). In the western part of its geographical range, *D. sagitta* is a habitat specialist occurring in sand habitats of the steppe, semi-desert and desert zones, whereas in the eastern part of the range, the species is a habitat generalist inhabiting different types of deserts (Shenbrot *et al.*, 1995). Based on the species geographical distribution and geographical variation in morphology and habitat selectivity, genetic differentiation should be more expressed in the western than in the eastern parts of its range and we would expect a cryptic speciation at the Kazakhstan/Central Asia boundary.

The aim of the present study was to examine the phylogeographical structure of *D. sagitta* with a focus on correlation between recognized morphological subspecies and genetic variation and to estimate the timing of divergence among the major geographical lineages.

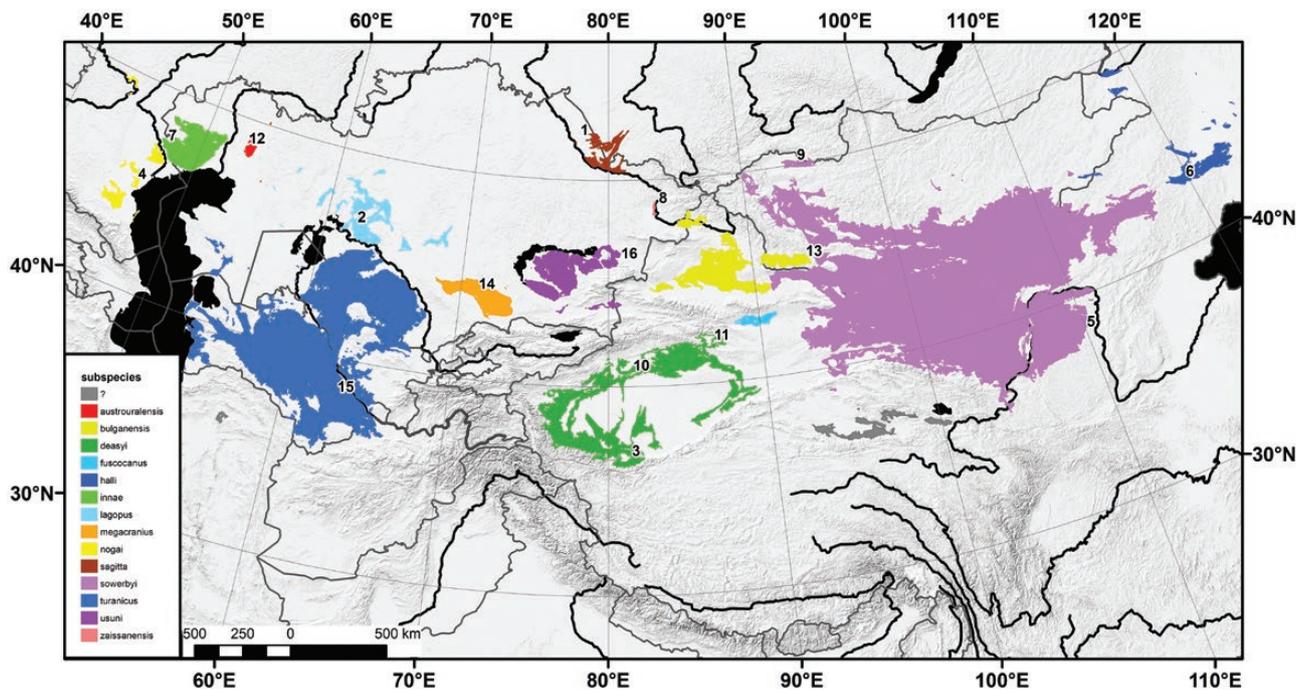


Figure 1. Map showing the species range and distribution of subspecies of *Dipus sagitta* following Shenbrot (1991a, b). Numbers denote the positions of type localities: 1 – *sagitta*, 2 – *lagopus*, 3 – *deasyi*, 4 – *nogai*, 5 – *sowerbyi*, 6 – *halli*, 7 – *innae*, 8 – *zaissanensis*, 9 – *ubsanensis*, 10 – *aksuensis*, 11 – *fuscocanus*, 12 – *austrouralensis*, 13 – *bulganensis*, 14 – *megaranius*, 15 – *turanicus*, 16 – *usuni*.

MATERIAL AND METHODS

SAMPLING, DNA EXTRACTION, AMPLIFICATION AND SEQUENCING

Most samples were obtained from fieldwork in Mongolia and Kazakhstan over the period from 2003 to 2016 by the Joint Russian-Mongolian Biological expedition and joint parasitology expedition of Pavlodar State Pedagogical Institute, Pavlodar University and Severtsov Institute of Ecology and Evolution RAN.

Mitochondrial DNA sequences were obtained from 182 individuals of *Dipus*, and 37 sequences were retrieved from GenBank (Supporting Information, Tables S1, S2). The total alignment of cytochrome *b* (*Cytb*) includes 219 sequences of *D. sagitta* from 73 localities across the species range (Figs 1, 3; Table S1) and three sequences of *Stylodipus* as the outgroup. Exon 11 of *BRCA1* and exon 1 of *IRBP* were examined in 73 and 74 *D. sagitta*, respectively, including two specimens from GenBank (Supporting Information, Table S2); two specimens of *Stylodipus* were used as the outgroup.

In most cases, total DNA was extracted from ethanol-preserved tissues (liver, kidney and toe clippings) using a standard protocol of proteinase K digestion, phenol–chloroform deproteinization and isopropanol precipitation (Sambrook, Fritsch & Maniatis, 1989). The entire *Cytb* gene (1140 bp) was amplified by polymerase chain reaction (PCR) with the forward/reverse primer combination L7-fw/H6-rev (Montgelard *et al.*, 2002) or L_glu_jak/H_thr_jak (Shenbrot *et al.*, 2017). Typical conditions for *Cytb* amplification included an initial denaturation at 94 °C for 3 min, 35 cycles of 94 °C for 30 s, annealing at 57–60 °C for 1 min and extension at 72 °C for 1 min, followed by a final extension at 72 °C for 10 min, and indefinite hold at 4 °C.

DNA of ten jerboas was extracted from dried skins of museum collection specimens (age range: 25–83 years), including the type specimen of *D. s. megacranius* Shenbrot, 1991 (ZMMU S-148042). In this case, DNA was purified directly using the MiniElute PCR Purification Kit (Qiagen) including an overnight lysis step following the manufacturer's protocol and recommendations of Yang *et al.* (1998). DNA extracted from six museum specimens was highly degraded, so only short fragments (140–280 bp) were obtained using the combination of internal primers designed for this study (Supporting Information, Table S3). The PCR programme for amplification of short fragments included an initial denaturation at 95 °C for 3 min, 45 cycles of 95 °C for 30 s, 48 °C for 30 s and 72 °C for 30 s, and a final extension of 72 °C for 6 min. All stages of the extraction process included a negative control run in parallel. To avoid contamination, extraction and amplification of the DNA from the museum specimens were carried out in the ZMMU Laboratory of Historical DNA, exclusively equipped for work with

museum DNA specimens, where no previous work on fresh tissues had been performed. We ran aliquots (10 µL) of the extractions alongside a 100-bp ladder on a 1% agarose gel by electrophoresis.

Exon 11 of *BRCA1* and exon 1 of *IRBP* were amplified and sequenced using external forward/reverse primer combinations as well as internal primers according to our previous studies (Lebedev *et al.*, 2012; Pisano *et al.*, 2015). PCR products were visualized on a 1% agarose gel and then purified using a Diaton DNA Clean-up kit. Approximately 30 ng of the purified PCR product was used for sequencing with each primer via an ABI 3100-Avant autosequencing system using ABI PRISM BigDye Terminator v. 3.1.

The sequences were deposited in GenBank under the following accession numbers: *Cytb*, MF535659–MF535796 and MF535902–MF535907; *BRCA1*, MF535849–MF535901; and *IRBP*, MF535797–MF535848.

MITOCHONDRIAL DATA

Tree reconstructions and distance estimation

Maximum likelihood (ML) reconstruction of the *Cytb* tree was performed in Treefinder, version October 2008 (Jobb, 2008). Appropriate models of sequence evolution were selected for each of the codon positions under the Bayesian information criterion (BIC) employing the routine implemented in Treefinder. Bootstrap analysis employed 1000 replicates.

A Bayesian tree reconstruction was conducted in MrBayes 3.2 (Ronquist *et al.*, 2012). Models with either two or six rate matrix parameters were selected for each partition based on the results of the model selection for the ML analysis. The analysis included two independent runs of four chains (one cold plus three heated following the default settings). The chain length was set to five million generations with sampling every 2000 generations. With these settings, the effective sample size exceeded 200 for all estimated parameters. Tracer 1.6 software (Rambaut & Drummond, 2005) was used to check for convergence and determine the necessary burn-in fraction, which was 10% of the chain length.

The ultrametric Bayesian tree was reconstructed in BEAST v. 1.82 (Drummond *et al.*, 2012) under the strict clock model. The choice of the model is validated by the results of the Likelihood Ratio Test (departure from the null hypothesis of rate constancy is non-significant: $P > 0.25$), which was conducted in PAML v. 4.7 (Yang, 2007) based on the ML topology. A piecewise constant coalescent skyline prior was used with ten groups. The partitioning scheme and models corresponded to those in the ML analysis. Two runs of 50 million generations were conducted. Parameter convergence was assessed in Tracer 1.6; burn-in was set to five million steps. The Maximum

Clade Credibility tree was generated by TreeAnnotator version 1.8.2 (part of the BEAST package).

Genetic distances among haplotypes were calculated through PAUP* 4.0b10 (Swofford, 2003). This program was also used to reconstruct the neighbor-joining (NJ) tree (uncorrected p-distance) for an extended alignment including short *Cytb* fragments obtained for museum specimens.

Species delimitation and spatial analysis

To define natural groups based on the mtDNA data, we employed two methods: the Automatic Barcode Gap Discovery (ABGD) method (Puillandre *et al.*, 2012) and the General Mixed Yule Coalescent (GMYC) model (Pons *et al.*, 2006; Fujisawa & Barraclough, 2013). Automatic identification of the 'barcode gap' was performed using the ABGD application available at <http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html> under the following parameters: P_{\min} (prior minimal distance) = 0.01, P_{\max} (prior maximal distance) = 0.1, X (relative gap width) = 1.0. A matrix of uncorrected p-distances was taken as input.

Single threshold GMYC analysis was performed employing the GMYC web server (<http://species.h-its.org/gmyc/>; Fujisawa & Barraclough, 2013) with the Maximum Clade Credibility tree produced by BEAST. Delimitation methods often tend to over-split taxa (Carstens *et al.*, 2013). This may be explained by insufficient complexity of the GMYC model, which treats all variation within a species as a coalescent within a single population. In a structured population, a coalescent can be divided into scattering and collecting phases (Wakeley, 1999) in which branch lengths are scaled differently, thus violating the assumptions of GMYC. To avoid over-splitting, we tried to reduce the contribution of intra-deme variation relative to inter-deme variation by using scattered sampling. The total sample was re-assembled so that each locality was represented by a single sequence selected at random. The GMYC analysis was repeated 50 times, and trees with a reduced number of tips were generated in Ape v. 3.4 (Paradis, Claude & Strimmer, 2004).

To examine the geographical variation at a finer scale, the Spatial Analysis of Molecular Variance (SAMOVA2; Dupanloup, Schneider & Excoffier, 2002) was performed with the samples of the most widespread lineages and the number of a priori defined groups (K) varying from 2 to 10.

NUCLEAR DATA

For allelic phase reconstruction, the PHASE module (Stephens, Smith & Donnelly, 2001, Stephens & Donnelly, 2003) implemented in the software DNAsp (v. 5; Librado & Rozas, 2009) was used. Alleles with

posterior probabilities below 0.9 were excluded from the analyses. Median-joining networks were reconstructed using Network version 5.0 under default options.

SPECIES TREE AND MOLECULAR DATING

To estimate the dates of divergence between the main lineages, we reconstructed the species tree employing a Bayesian coalescent framework as implemented in *BEAST v. 1.8.2 (Heled & Drummond, 2010). This method implies that gene tree/species tree conflicts are accounted for by incomplete lineage sorting but not by gene introgression. Therefore, lineages for which cytonuclear discordance was suspected were excluded from the analysis. *Stylodipus* was used as the outgroup.

The faster evolving mitochondrial genes are subject to rate decay due to saturation and non-neutral variation. Evidence for saturation was examined by plotting uncorrected pairwise differences (p-distances) for different substitution classes/codon positions against ML distances for the complete gene. To mitigate the effect of saturation, only third codon position transversions were retained in the analysis. Separate strict clock models were used for each of the three loci. *IRBP* and *BRCA1* were partitioned into two subsets each (1st+2nd and 3rd codon positions), the optimum models for which were selected in Treefinder. 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. The tree was calibrated based on the earliest fossil *Dipus* (Shala Formation, middle Baodean, China; Qiu, Wang & Li, 2006). The minimum age of this fossil (8.2 Myr) was used as the hard lower bound for the time of its divergence from *Stylodipus*. Prior calibration density was modelled using an exponential distribution. Its parameter was chosen so that the soft (95%) upper bound corresponded to 10 Myr. Two runs of 100 million generations were conducted. Parameter convergence was assessed in Tracer.

To infer the timing of population events such as demographic changes, only mitochondrial data were used. However, the phylogenetic rate of the *Cytb* gene should not be used for this because of the time dependency of rate estimates (Ho *et al.*, 2007). To obtain a rate estimate applicable for recent events, we used the procedure assuming that the rate of the third position transversions (μ_{tv3}) is negligibly affected by rate decay, and hence the phylogenetic rate of change for this substitution class can be used for calibration in the historical demographic analysis. The total number of nucleotide substitutions observed within the population-level clusters and, separately, of the third position transversions was calculated in PAUP* 4.0b10. The ratio of these two numbers divided by three and

multiplied by the estimate of μ_{tv3} from *BEAST was used as the population-level rate of change for the whole gene. Its approximate standard error was calculated using the delta-method.

NEUTRALITY TESTS AND ANALYSES OF DEMOGRAPHIC HISTORY

We tested the hypothesis of demographic stability by calculating Fu's neutrality statistic F_s (Fu, 1997) and Tajima's D test (Tajima, 1989) in Arlequin 3.5. The significance of the statistics was tested by generating 1000 random samples under neutrality. Demographic histories of the main *Cytb* phylogroups were inferred by pair-wise mismatch distribution analyses (Rogers & Harpending, 1992) as implemented in Arlequin. The validity of the stepwise expansion model was tested using parametric bootstrapping with 1000 replicates. Skyline plots for the best-sampled mtDNA lineages were generated in BEAST. The *Cytb* alignment was partitioned into two subsets including 1st+2nd and 3rd codon positions, respectively. The analysis was run for 100 million steps under a strict clock model using the skyline prior with ten groups. The times of expansion were estimated based on tau parameters of the demographic expansion model using the population-level rate calculated as explained above and assuming a generation time of 1 year.

RESULTS

CHARACTERISTICS OF *CYTB* SEQUENCE

Among 219 *Dipus* specimens, 164 unique haplotypes of *Cytb* were found (Arlequin 3.5 results). The length of sequences was complete or nearly complete (1125–1140 bp) for most, including a museum specimen from Semipalatinsk (type locality of *D. sagitta*). Shorter sequences (266–723 bp) were obtained for nine other museum exemplars. Sequences less than 500 bp in length were used only in the NJ analysis.

The number of nucleotide substitutions observed among sequences of *D. sagitta* were as follows: third positions – 331 (including 48 transversions), first and second positions – 61 substitutions; 40 substitutions were non-synonymous.

Saturation plots (Supporting Information, Fig. S1) demonstrate that the third position transitions are saturated for comparisons among the most divergent lineages of *D. sagitta*.

CYTB PHYLOGENETIC TREE

The ML phylogenetic analysis of the *Cytb* data set revealed six highly supported main lineages, some of which, in their turn, are structured at a finer scale

(Figs 2, 3; see also the NJ tree: Supporting Information, Fig. S2). Justification of this structuring is given below under 'Species delimitation'. The level of divergence between the main lineages was between 7.4 and 10.1% (average uncorrected distance between haplotypes).

Lineage I: Jungaar Basin (Dzungaria) and NE Kazakhstan. This is subdivided into three sublineages separated by a p-distance of 3.4–4.0%: Ia – Irtysh valley (type locality of *D. sagitta*, represented by a single specimen); Ib – Jungaar Basin, Mongolian Dzungaria (Baruun-Hure) and Kara-Irtysh valley; Ic – western part of the Zaisan depression.

Lineage II: most of Gobi, Alashan and adjacent deserts of Mongolia and China. This includes several sublineages divergent by 2.2–3.4%: IIa – most of the Mongolian range (including Transaltai Gobi, Mongolian South Gobi and part of East Gobi), Alashan, N Gansu, NE Qinghai; a subcluster (designated IIa₀) is restricted to the Great Lake depression and the Valley of Lakes in Mongolia; it is divergent from the rest of IIa by 1.3%; IIb – most of East Gobi and isolates in central Mongolia, sympatric with Ia at one site in East Gobi; IIc – dominant in Ordos, Ningxia (sympatric with IIa); IID – east-central Nei Mongol.

Lineage III: NW Mongolia and Tuva (Uvs and Achitnur depressions, part of the Great Lake depression); the haplotype from the Uvs-nuur depression (IIIb) is quite divergent (2.3%) from the others (IIIa).

Lineage IV: SE and C Kazakhstan – southern part of the Balkhash depression, Ili valley, Muyunkum sands, North Aral region.

Lineage V: western part of the range, and includes two sublineages differing by 4.7%: Va – W Turan (Qyzylkum, S Turkmenistan, NW Kazakhstan); Vb – Dagestan.

Lineage VI: western and southern parts of the Qaidam Basin.

In all cases, haplotypes belonging to different main lineages are distributed allopatrically (Fig. 3). Likewise, sublineages within the main clades are mostly restricted to non-intersecting fragments of the range but show few instances of sympatry/parapatry.

The phylogenetic relationships among lineages are not fully clear. Lineage VI is placed as sister to all other clades with moderate support; the relationships among the other clades appear unresolved.

SPECIES DELIMITATION

ABGD method

The distribution of genetic distances between haplotypes (Supporting Information, Fig. S3) has three peaks with the more pronounced gap corresponding to ~5–6%;

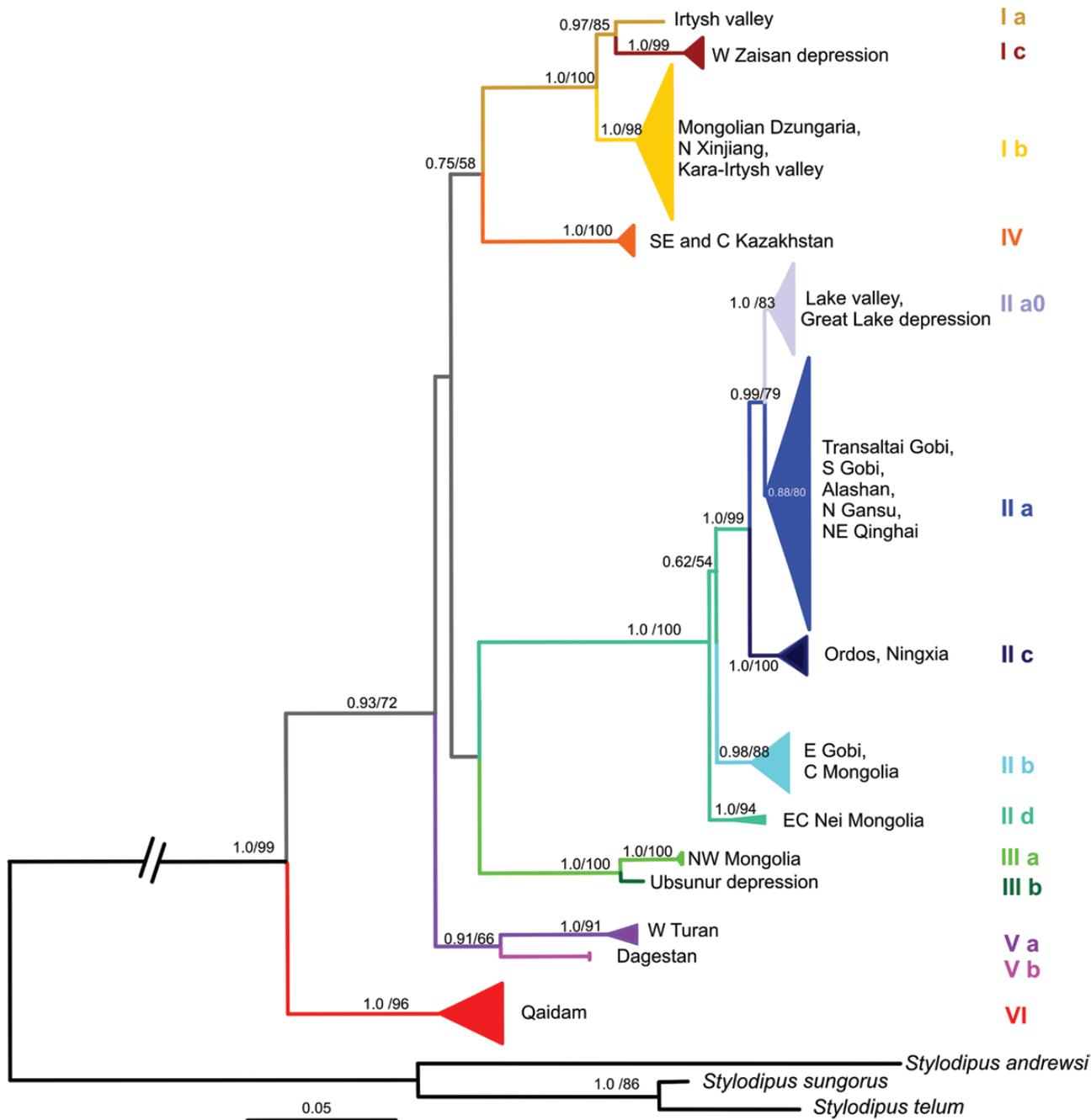


Figure 2. The maximum likelihood (ML) phylogeny of *Dipus sagitta* as inferred from the *Cytb* data from 222 specimens including three species of *Stylodipus* as the outgroup. Numbers above/below branches correspond, respectively, to the Bayesian posterior probabilities (MrBayes) and ML bootstrap support (>50%) for the main clades.

accepting 5.5% as the cut-off threshold, the tree is divided into six main lineages (I–VI). This subdivision pattern is inferred by ABGD if the prior for maximum intraspecific diversity (P) is set to values $> 1\%$. With $P < 0.8\%$, the sample is partitioned into 13 subsets (Ia–Ic, IIa + IIc, IIb, II d, IIIa, IIIb, IV, Va, Vb and two sublineages within VI).

GMYC model

With the total unmodified sample, GMYC definitely over-splits the data, suggesting the existence of 28 entities. Under scattered sampling, the number of inferred groups decreased dramatically. In 39 of 50 replicates, 13 entities were recognized. They correspond

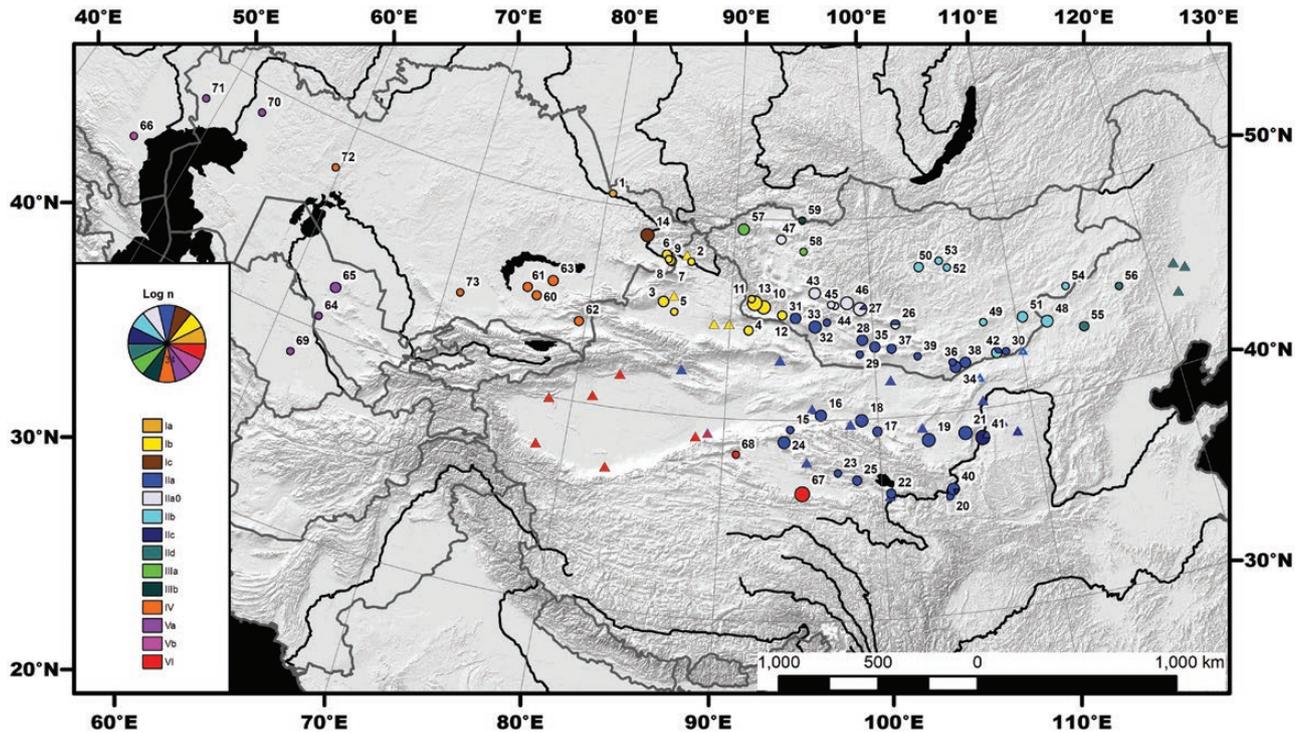


Figure 3. The geographical distribution of sampling localities and *Cytb* phylogroups. Symbols are coloured by mitochondrial lineage as in Figure 2. Circles indicate sampling localities examined in this study. Locality codes and detailed geographical information are given in Supporting Information, Table S1. Triangles denote sites sampled by Cheng *et al.* (2017).

to lineages Ia–Ic, Ila–IId, IIIa, IIIb, IV, Va, Vb and VI. Thus, the suggested modification of the GMYC procedure allowed us to arrive at a more reasonable solution.

SAMOVA RESULTS OF *CYTB* DATA

The results of the SAMOVA analysis of lineage I show a clear peak in both Va (among-group variance component) and Fct (proportion of variation among groups) corresponding to a number of groups (K) = 3. This subdivision is consistent with partitioning into sublineages Ia, Ib and Ic. The analysis of molecular variance demonstrates that 81% of the overall population variation is explained by the divergence among these three groups. Differentiation was stronger within populations (individual variation) than among populations within groups (4 and 15% of the overall population variation, respectively).

The results for lineage II were more ambiguous: Fct increased from K = 2 to 9 with the highest delta Fct at K = 5, while Va had a global maximum at K = 3 and a local maximum at K = 6. At K = 3, the data are subdivided into subsets corresponding to sublineages IIIb, IId and Ila + Iic. The last cluster also includes locality 42, which is notable for co-occurrence of Ila and Iib. Differentiation among groups, among populations

within groups and within populations accounts for 61, 18 and 21% of the total variation, respectively. The solution with K = 5 is coherent with partitioning into sublineages Iib, Iic, Iid, Ila₀ and Ila without Ila₀.

Thus, the SAMOVA results indicate that the pattern of geographical subdivision within lineages is concordant with the GMYC results.

NUCLEAR DATA

BRCA1

Based on the sample of 74 individual sequences (834 bp), PHASE reconstructed 38 haplotypes. Observed heterozygosity was 0.5 (37 heterozygotes). The allelic phase could not be determined with adequate posterior probability (>0.9) in seven sequences, which were excluded from the analysis. The relationships among the remaining 28 haplotypes as inferred by NETWORK and Treefinder are shown in Figure 4A and Supporting Information, Fig. S4, respectively.

Within this sample, 37 nucleotide substitutions and 21 amino acid substitutions are observed. The consistency index for the nucleotide data is 0.88. In the network we identified clusters of haplotypes corresponding to mtDNA lineages Ib + Ic, IV + V and VI.

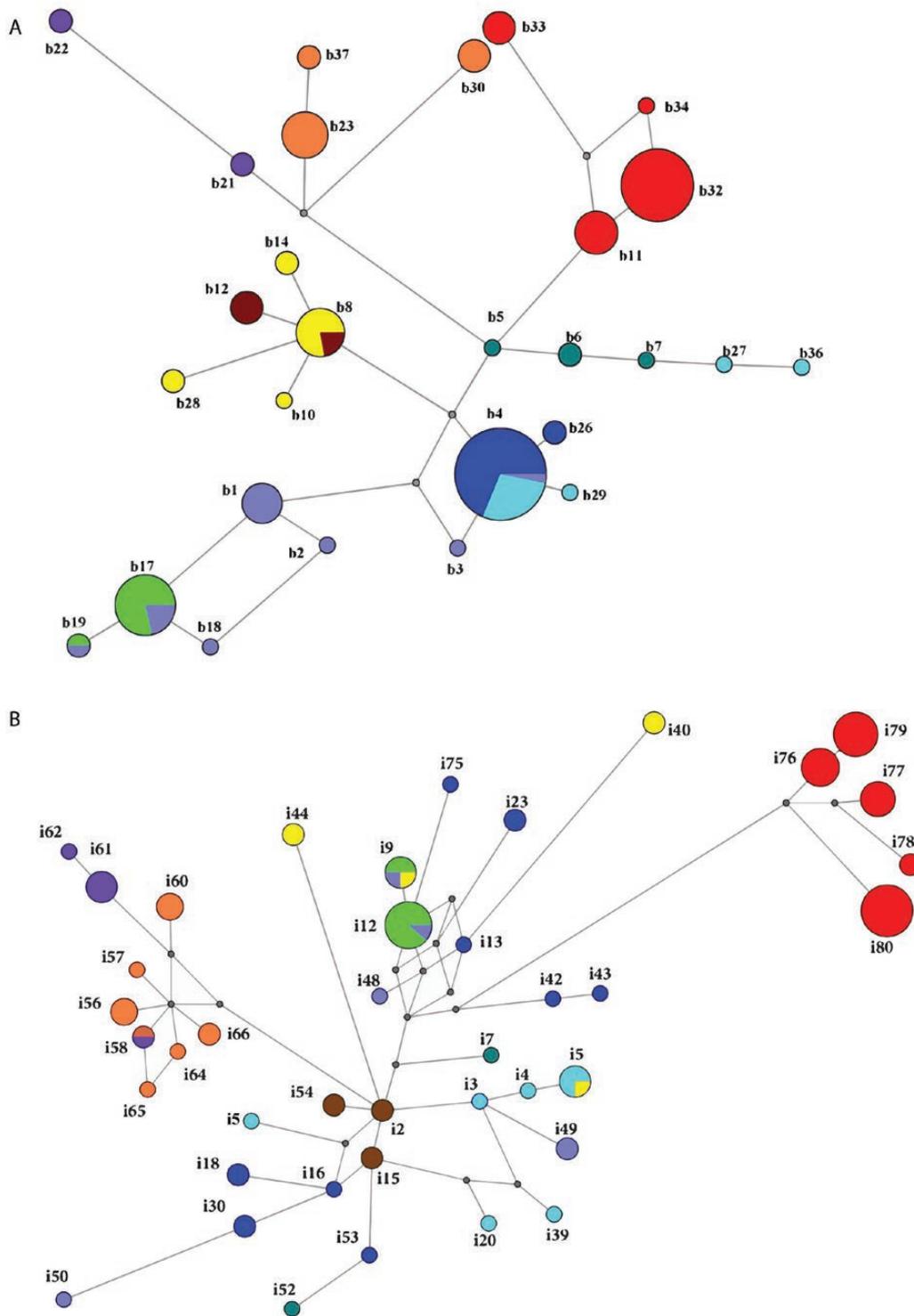


Figure 4. Median-joining network showing the relationships among the alleles of A, *BRCA1* and B, *IRBP* in *Dipus sagitta*. Allele and specimen codes are given in Supporting Information, Table S1. The size of circles corresponds to the number of specimens with identical alleles. The colours denote corresponding mitochondrial lineages (see Figs 2, 3).

The relationships among haplotypes from Gobi and NW Mongolia (mitochondrial lineages II and III) are less clear. In particular, most of sequences of exemplars

from the Valley of Lakes and Great Lake Depression (mitochondrial sublineage IIa₀) cluster together with those from NW Mongolia (mitochondrial lineage III).

IRBP

Based on the sample of 76 individual sequences (1035 bp), PHASE reconstructed 80 haplotypes. Observed heterozygosity was 0.8625 (69 heterozygotes). The allelic phase was determined with posterior probability of >0.9 in 50 sequences. The relationships among the 41 haplotypes as inferred by NETWORK are shown in Figure 4B. The ML tree is presented in Supporting Information, Fig. S5.

Within this sample, 60 nucleotide substitutions and seven amino acid substitutions are observed. The consistency index for the nucleotide data is 0.645. As with *BRCA1* results, the network contains divergent clusters corresponding to mtDNA lineages IV + V and VI. No evident structure among haplotypes from Mongolian, Kazakh and most of Chinese parts of the range (mitochondrial lineages II and III) is found. This may be partly associated with a high level of homoplasy due to parallel mutations in third codon positions, which are typical for this CG-rich marker (CG content of 77.1% vs. 40.6% in *BRCA1*). Again, animals from NW Mongolia and the Valley of Lakes (lineages III and IIa₀) share common alleles, suggesting past or recent gene flow.

The direct comparison of nuclear networks and the *Cytb* tree suggests that at least some of the mitochondrial lineages correspond to 'true' taxa that are also supported by the nuclear data. Both genes support lineage VI and separate lineages IV and V from the rest. Lineage I is supported by *BRCA1* but not by *IRBP*. Therefore, we consider it appropriate to include both nuclear and mitochondrial data for the majority of lineages in a species tree reconstruction. Taking into account potential cytonuclear discordance between lineages II and III, the latter lineage and sublineage IIa₀ were omitted.

SPECIES TREE, MOLECULAR DATES AND SUBSTITUTION RATES

The set of taxa ('species') included Dzungarian (Ib), Zaisan (Ic), Gobi (IIa without IIa₀), W Turan (IVa), SE Kazakhstan (V) and Qaidam (VI). The species tree reconstructed by *BEAST (Fig. 5) strongly supports the position of the Qaidam lineage, which branches from the root node at ~ 1.5 Myr [95% highest probability density (HPD): 0.97–2.2 Myr, Early Pleistocene]. A well-supported clade comprising W Turan and SE Kazakhstan lineages branches off at ~ 0.9 Myr (the end of the Early–early Middle Pleistocene, 95% HPD: 0.5–1.4 Myr), and the age of the split between these two lineages dates back to ~370 kya (Middle Pleistocene, 95% HPD: 80–730 kya). The Zaisan + Dzungarian and Gobi lineages are recovered as sister groups diverging at ~0.46 Myr (Middle Pleistocene, 95% HPD: 240–760

kya). The two branches within the former split at the end of the Middle Pleistocene (~0.25 Myr, 95% HPD: 70–470 kya).

The *BEAST analysis produced the following estimated substitution rates of the genes employed (/site/Myr): 0.24% – *BRCA1*, 0.53% – *IRBP*, 1.44% – third position transversions of *Cytb*. Based on the last value, we estimated the rate of *Cytb* appropriate for population-level analysis as 7.3%/site/Myr (SE = 1.9%).

Comparison between the species tree and the ultrametric tree inferred from the *Cytb* alignment (Fig. 6) highlights potential disagreement between the mitochondrial and nuclear data. Accepting the inferred *Cytb* rate of 7.3%, the expected time of divergence of the Qaidam lineage is ~1.24 Myr, while the species tree estimate is 20% older. The most realistic explanation is that, at this level of divergence, there is some rate decay in the mtDNA clock, which is probably accounted for by saturation. Besides, mtDNA suggests that the divergence of the main lineages (except the Qaidam lineage) occurred within a relatively narrow time interval (at ~0.7–1.0 Myr); by contrast, in the species tree, the split between the W Turan + SE Kazakhstan and Dzungarian + Gobi lineages takes place at a much earlier stage than the divergence within these clades. However, we treat the latter result as preliminary and suggest that it should be tested with a larger sample of nuclear loci.

Note that our date estimates are significantly younger than those inferred in a recent study on Chinese *Dipus* (Cheng *et al.*, 2017); however, the latter are based on an a priori assumption that the *Cytb* rate is close to 2.0%, which is not supported by fossil evidence.

HISTORICAL DEMOGRAPHY

For most of the lineages and populations examined, the neutrality tests revealed no tendency for population growth. However, for lineage IIa (most of the Gobi range), Fu's *F*_s and Tajima's *D* tests were both significantly negative, suggesting population expansion (Table 1; Fig. 7A, B, D). The same result was obtained for sublineage IIa₀ (Great Lake depression and the Valley of Lakes in Mongolia) and local samples from southern and south-western Mongolia. Tests based on sum of square deviations and raggedness index did not reject the sudden-expansion model. In agreement with the above, skyline plots showed recent rapid population growth in lineage IIa but no significant departure from constancy in lineage Ib (Fig. 7C, E). Applying the *Cytb* rate of 7.3% to the skyline results, we estimate the start of expansion in IIa (without IIa₀) at ~19 kya. The values of tau parameters of

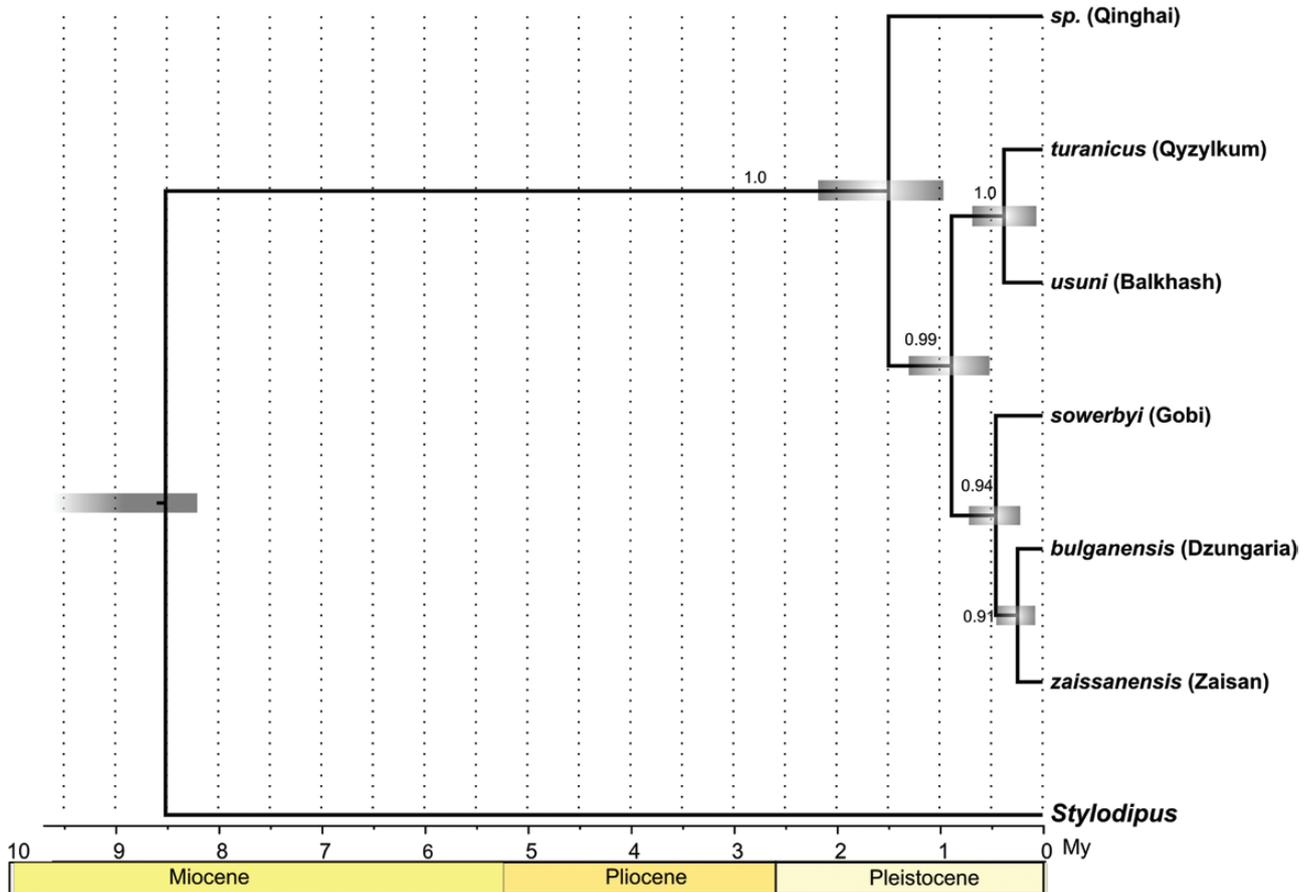


Figure 5. Timescale of major divergence events among *Dipus sagitta* based on the species tree produced by the *BEAST algorithm. The divergence times correspond to the median posterior estimates of their age in Myr. The grey bars represent the 95% highest probability density intervals. Values above the branches correspond to posterior probabilities for each node.

sudden expansion translate to 7–23 kya for the South and Transaltai Gobi populations and ~30 kya for the Lake valley lineage.

DISCUSSION

EVOLUTIONARY AND DEMOGRAPHIC HISTORY

Dipus sagitta provides another example of deep genetic subdivision within a widely distributed Palearctic desert rodent. Similar cases of cryptic or incipient species (semispecies) are presented by *J. jaculus* (Ben Faleh *et al.*, 2012b), *Paralactaga elater* (Mohammadi *et al.*, 2016) and *M. meridianus* (Neronov *et al.*, 2009; Ito *et al.*, 2010).

Most *Dipus* fossils including the earliest (*Dipus cf. sagitta*) are found in the late Neogene and Pleistocene of China and Mongolia (Zazhigin & Lopatin, 2001; Li & Qui, 2005) while all fossils from the western part

of the range are of Late Pleistocene or Holocene age (Zazhigin & Lopatin, 2001). This suggests that the primary centre of origin of the *D. sagitta* complex lies in the eastern part of its contemporary range. The fact that East Central Asian deserts (Gobi, Tarim Basin, Alashan, Jungaar Basin) harbour several unrelated *Dipus* lineages, including the most divergent one, lends support to this hypothesis.

Regardless of the true pattern of relationships among major lineages, the bush-like appearance of the mtDNA tree suggests that there was a burst of radiation in the late Early–early Middle Pleistocene (0.7–1.0 Myr). This rapid branching may have been associated with a range expansion and, in particular, with the dispersal to Turan (West Central Asian deserts). Given the lack of monophyly of western clades (I, IV and V), the areas west of Tian-Shan and Altai could have been colonized independently several times by different lineages, and perhaps via

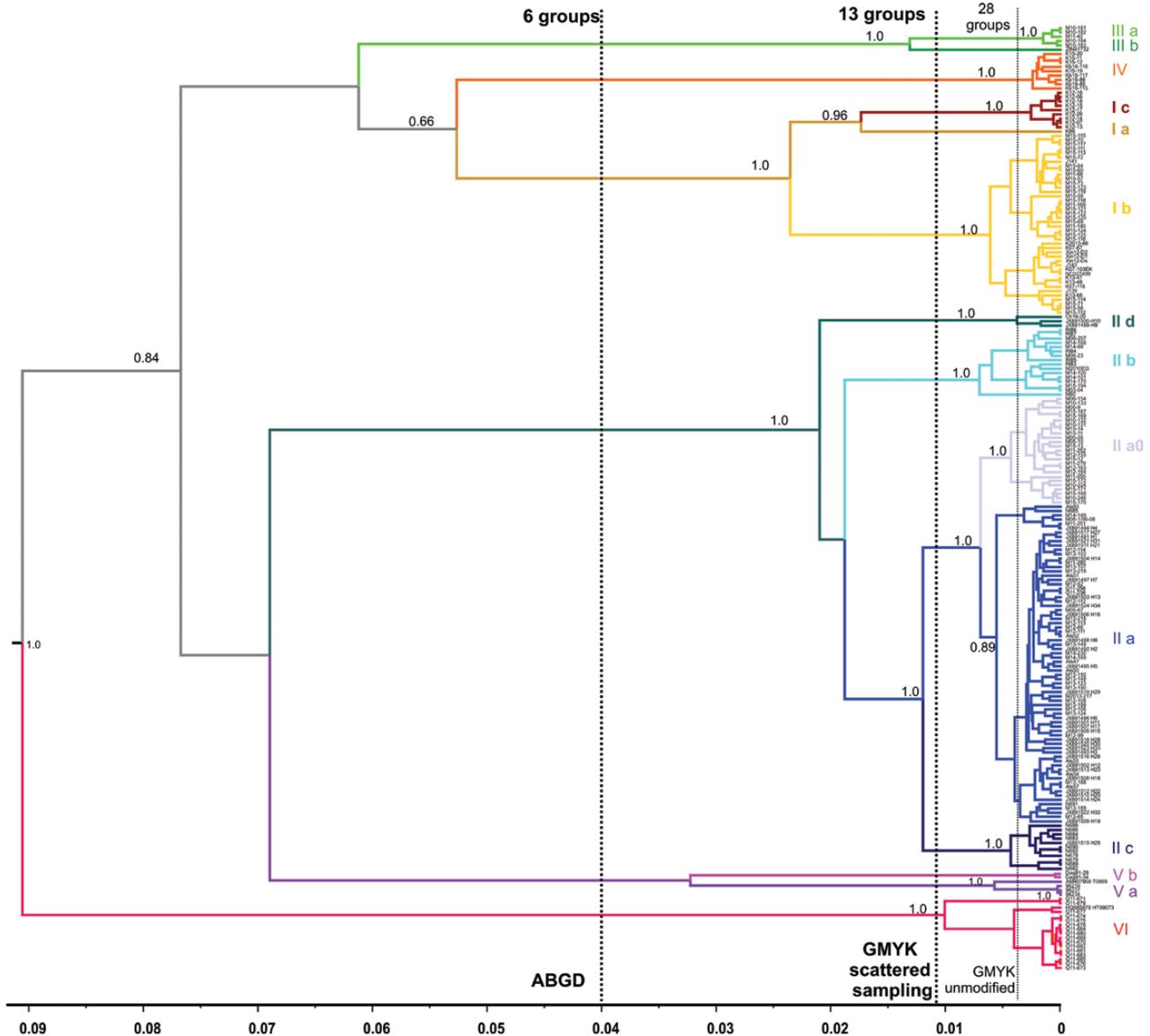


Figure 6. The ultrametric Maximum Clade Credibility (MCC) tree produced by BEAST based on the *Cytb* alignment. The results of species delimitation using ABGD (Automatic Barcode Gap Discovery) and GMYC (General Mixed Yule Coalescent) methods (vertical dotted lines) show the number of subsets identified by each method. The horizontal axis represents time in substitutions per site.

different routes. It remains to be clarified whether this diversification shift can be accounted for by a major climate transition towards more arid conditions associated with the establishment of the 100-kyr periodicity of ice ages (Head & Gibbard, 2015) or other Early–Middle Pleistocene events (Han, Fang & Berger, 2012).

There are clear ecological differences between western and eastern groups of populations of *Dipus* (Shenbrot *et al.*, 1995): while most western populations are strictly psammophilous, in the eastern part of the range *Dipus*

is a habitat generalist, and only near the north-eastern limits of its distribution does it again become a habitat specialist occurring in sand ‘islands’ of the steppe zone. The factors responsible for the increase in habitat specialization during westward expansion of *Dipus* are not clear. Generally, habitat specialization is better expressed in desert rodents of Turan and Kazakhstan than in those of China and Mongolia (Shenbrot *et al.*, 1999). The increased specialization of the western populations of *Dipus* is probably a result of greater interspecific competition in the west Central Asian deserts. One

Table 1. Results of the neutrality tests and mismatch distribution analysis

Lineage	Population	Sample size	Nucleotide diversity (per locus)	Tajima's <i>D</i>	Tajima's <i>D P</i> value	<i>F_s</i>	<i>F_s</i> <i>P</i> value	Tau	Expansion time (kya)	
Ib	Dzungaria	42	8.8	-0.80	0.22	-5.45	0.048	11.24	67.5	
	Mongol Jungaar	28	7.6	0.21	0.64	-0.94	0.38	11.38	68.3	
	Kara-Irtysh valley	7	5.9	-1.10	0.17	0.62	0.57	2.88	17.3	
Ic	Zaisan depression (west)	9	4.6	1.11	0.88	0.99	0.68	5.62	33.7	
IIa	Gobi, Alashan, Lake valley	68	10.5	-1.69	0.023	-24.53	0.000	13.97	83.9	
	Valley of Lakes	23	7.4	-1.46	0.051	-10.47	0.000	7.86	47.2	
	Transaltai and South Gobi	27	5.8	-2.38	0.000	-18.51	0.000	1.83	11.0	
	NW Transaltai Gobi	9	3.6	-1.67	0.024	-3.71	0.015	1.13	6.8	
	E Transaltai Gobi	8	3.4	-1.60	0.033	-5.13	0.001	3.88	23.3	
	S Gobi	8	5.8	-1.49	0.050	-3.50	0.023	2.68	16.1	
	Alxa	8	7.0	-1.11	0.14	-3.00	0.038	6.60	39.6	
	IIb	East Gobi, Central Mongolia	15	10.1	-0.93	0.17	-1.96	0.17	10.55	63.3
		East Gobi (IIb + admixture of IIa)	13	17.7	-0.87	0.17	-0.52	0.36	0.09	0.5
IIc	Ningxia (IIc + admixture of IIa)	12	13.9	-0.73	0.22	-0.62	0.33	7.27	43.6	
V	Balkhash depression	9	4.83	-1.11	0.16	-2.82	0.045	7.95	47.7	
VI	Qinghai (Golmud)	16	6.43	-0.98	0.19	6.02	0.98	2.37	14.2	

Expansion times were calculated using the *Cytb* substitution rate of 7.3%/Myr/site. Data for populations showing significant expansion are given in bold.

might expect that, due to habitat restriction and range fragmentation, the western group of populations would be more differentiated genetically than the eastern one. However, rigorous testing of this hypothesis is complicated by the fact that the eastern group is highly variable due to its older age.

At the intraspecific level, rodents of the arid zone demonstrate shallow geographical genetic structure in most cases examined (Ben Faleh *et al.*, 2012b; Lv *et al.*, 2016; Ndiaye *et al.*, 2016). The only known exception is *Jaculus orientalis* (Ben Faleh *et al.*, 2012a), but this species is not a true desert form, occurring in Mediterranean semi-deserts. Compared to other taxa, *Dipus* is remarkable for its complex nested structure with several major lineages subdivided at a lower level into a number of sublineages (lineages I and II). In lineage II, no major barriers responsible for geographical differentiation are evident. This indicates that the idea of weaker structuring in desert vs. steppe and forest species is not universally supported and should be tested with more data. We hypothesize that the pattern of variation observed in *Dipus* may be the result of historical persistence and/or survival in multiple refugia combined with relatively slow dispersal from these refugia.

In most of the populations examined, the analysis of demographic history revealed no significant changes in effective population size. Presuming that cold phase conditions of the last glacial cycles are not unfavourable for desert rodents, this result is not unexpected. However, the populations of the extra-arid parts of the Gobi (Transaltai Gobi, South Gobi) show a clear signal of recent expansion (this study; Liao *et al.*, 2016; Cheng *et al.*, 2017). Our dating results are consistent with post-glacial expansion, suggesting that the last glacial maximum climate change rendered extra-arid habitats unsuitable even for species that are well adapted to it.

GENETIC LINEAGES AND MORPHOLOGICAL SUBSPECIES OF *D. SAGITTA*

Both mitochondrial and nuclear data indicate significant differentiation within *Dipus*, which needs to be reviewed from a taxonomic perspective.

Previous revision of *Dipus* (Shenbrot, 1991a, b) suggested recognition of 12 subspecies combined into two subspecies groups, an eastern 'sagitta' group with five subspecies (*D. s. sagitta* Pallas, 1773, *D. s. zaisanensis* Selevin, 1934, *D. s. sowerbyi* Thomas, 1908, *D. s. halli* Sowerby, 1920 and *D. s. bulganensis* Shenbrot, 1991)

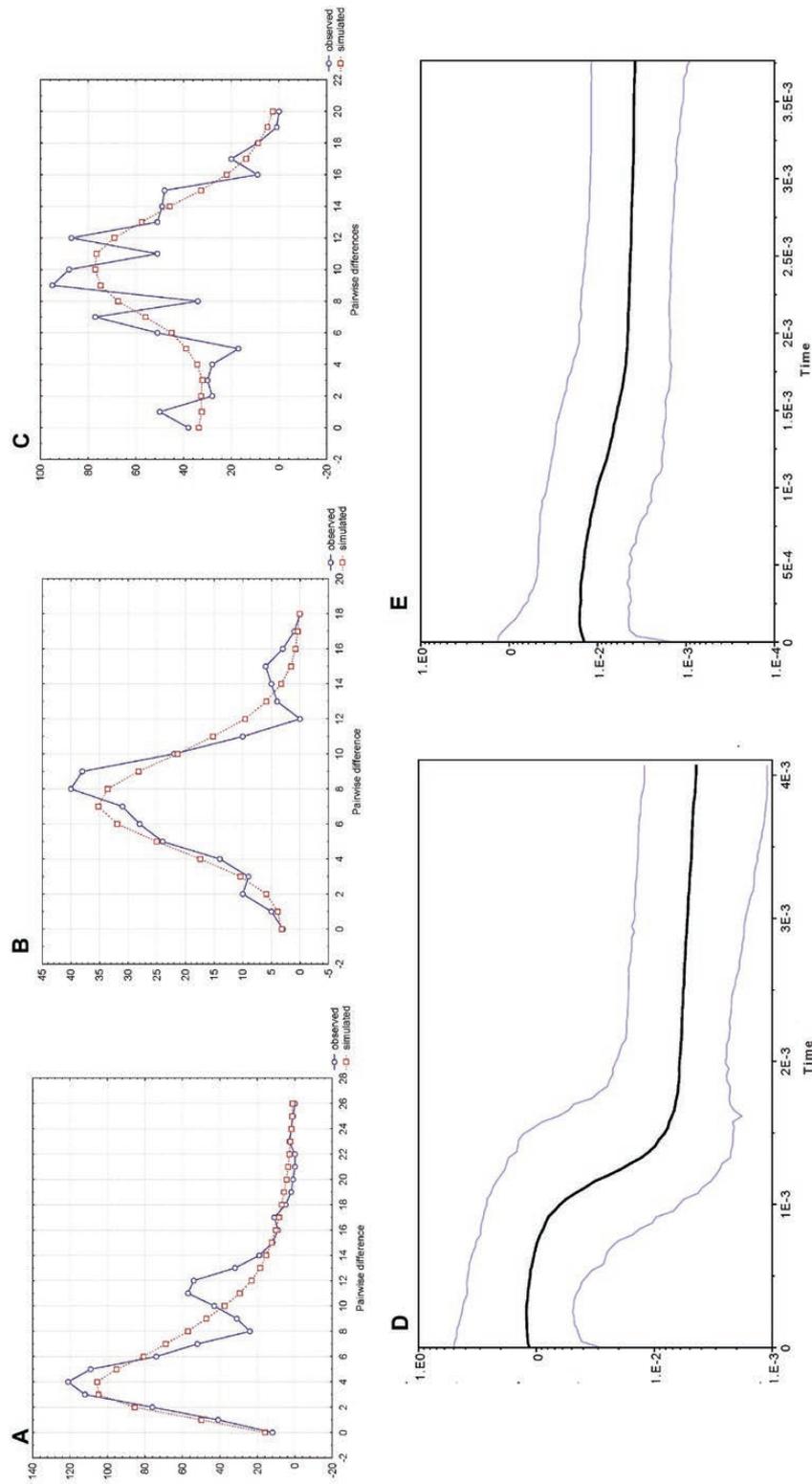


Figure 7. The results of historical demographic analyses (*Cytb* gene) of the widespread lineages Ib and IIa: A–C, mismatch distributions in lineages IIa (without IIa₀), IIa₀ and Ib, respectively; D, E, the Bayesian skyline plots for lineages IIa and Ib, respectively.

and a western ‘*lagopus*’ group with seven subspecies (*D. s. lagopus* Lichtenstein, 1823, *D. s. nogai* Satunin, 1907, *D. s. innae* Ognev, 1930, *D. s. austrouralensis* Shenbrot, 1991, *D. s. megacranius* Shenbrot, 1991, *D. s. turanicus* Shenbrot, 1991 and *D. s. usuni* Shenbrot, 1991). These two subspecies groups are clearly distinguishable based on the morphology of foot and glans penis and by chromosome complement ($2n = 48$ in both groups, whereas autosome fundamental number = 90 in the western group and 92 in the eastern group) (Shenbrot *et al.*, 1995). Within each group, the subspecies differ distinctly in size and proportions of the skull. However, the revised material did not include jerboas from the Qaidam and Tarim Basins (*D. s. deasyi* Barret-Hamilton, 1900, *D. s. aksuensis* Wang, 1964 and *D. s. fuscocanus* Wang, 1964).

Dipus sp. from Qaidam and Tarim Basins

The main result of our molecular study is the existence of a highly divergent lineage distributed in Qaidam, in particular in its southern and north-western parts. A separate position of this lineage is supported by both mitochondrial and nuclear data which suggest that it split from the rest of *Dipus* in the Early Pleistocene (at ~1.5 Myr and not later than 950 kya). Moreover, it is noteworthy that the north-east Qaidam basin is also populated by a different lineage that is widespread across South Mongolia and northern China. Although the two lineages are found within the same region, the available genetic data provide no indication of gene flow between these two taxa. Based on the above it seems reasonable to conclude that the Qaidam lineage should be regarded as a separate species.

The proper name of this species requires separate consideration. Neither taxon of *Dipus* has its terra typica in Qaidam. However, it is evident that the range for the species is not restricted to the latter area. A recent study of genetic variation in Chinese populations of *Dipus* (Cheng *et al.*, 2017) showed that mitochondrial lineage VI is also distributed throughout the Tarim Basin. The latter area harbours type localities of *D. s. deasyi* Barret-Hamilton, 1900 (Nura, southern Tarim Basin) and *D. s. aksuensis* Wang, 1964 (Tza-mutai, near Aksu, northern Tarim Basin).

Given that the two names apparently refer to the same taxon, we believe that *deasyi* Barret-Hamilton, 1900 is the valid name for the newly recognized species.

The available morphological diagnosis of *D. deasyi* is uninformative and requires revision. The original description indicates that its ‘teeth are more massive’ than in *D. lagopus* (= *sagitta*). Later, Thomas (1908) remarked that *deasyi* can be distinguished from easterly *sowerbyi* by smaller bullae and more slender rostrum. Our preliminary morphological data indicate

that molars of jerboas from southern Qaidam are relatively wider and have a more complicated structure than other contemporary forms of *Dipus* due to better development of ento- and ecto-lophs and lophids. We examined the skulls of two specimens (collection of the Zoological Institute of RAS; N 2757, 2255) collected near Hotan (southern Tarim Basin), which are expected to belong to *D. s. deasyi*, and found that, in their molar morphology, they are similar to animals from Qaidam. However, further work is required to reveal diagnostic traits that can be effectively used for discrimination between *D. deasyi* and other taxa.

Genetic lineages I–V

Concerning the other genetic lineages, there are corresponding available names for most of them.

The nominotypical subspecies (terra typica in the Middle Irtysh) is represented in our sample by a single specimen belonging to lineage Ia.

Lineages Ib and Ic correspond to the subspecies *D. s. bulganensis* (type locality is near Loc. 12) and *D. s. zaissanensis*, respectively. The latter appears to be restricted exclusively to near its type locality (sands of the north-western part of the Zaisan depression, Loc. 6). The eastern part of the Zaisan Basin is populated by a more widespread *D. s. bulganensis*, the two forms being separated there by ~150 km with no significant barrier. However, there is no indication of gene flow between them.

Jerboas from Gobi, Alashan and Ordos that are attributed to the morphologically distinct *D. s. sowerbyi* belong to mitochondrial lineage II, which includes several sublineages. Based on the results of Cheng *et al.* (2017), we might expect that the type locality (Yulin-fu, Ordos, Shaansi) is dominated by the lineage IIc. However, the status of the latter is uncertain; in Ordos it is sympatric with a more widespread lineage IIa and it remains unclear whether these two lineages correspond to different subspecies. Geographically close populations of *D. s. sowerbyi* from the north-western Transaltai Gobi and *D. s. bulganensis* from Mongolian Dzungaria show no signature of gene exchange; this highlights the importance of a barrier separating Mongolian and Dzungarian fauna.

The status of the easternmost subspecies *D. s. halli* has been in dispute. Based on its larger size, it was suggested that it is a separate species (Fomin & Lobachev, 1988). However, Shenbrot (1991b) did not accept this treatment. The range of *D. s. halli* is also a matter of controversy. Shenbrot believed that this subspecies is distributed only in east-central Inner Mongolia (Xilinggol, Tongliao, Chifeng) and south-easternmost Mongolia (Dariganga) while Fomin & Lobachev (1988) suggested a wider range including East Gobi. Our data provide no support for the species status of *D. s. halli*,

which probably corresponds to lineage II_d. The status of lineage II_b (East Gobi, isolates of central Mongolia) remains to be established. The fact that ‘sand island’ populations of central Mongolia, which are separated from the main range by more than 200 km, share the mtDNA haplogroup with jerboas from East Gobi indicates ancient northward expansion with an origin in the latter area. The age of this expansion can be tentatively estimated at ~60 kya.

The existence in NW Mongolia of a highly divergent lineage (III) is an unexpected result of this study. This area includes the terra typica of *D. s. ubsanensis*, which morphologically is similar to *D. s. sowerbyi* and was included within it by Shenbrot (1991b). It seems premature to make any conclusions on the status of *D. s. ubsanensis*, because the relationships between the latter and *D. s. sowerbyi* are not fully clear. In particular, the available nuclear data indicate that populations from the Valley of Lakes and Great Lakes Basin (mtDNA lineage II_a) are genetically closer to *D. s. ubsanensis* than to *D. s. sowerbyi*. Mitochondrial similarity between Gobi and Valley of Lakes populations can then be explained by introgression of alien haplotypes through secondary contact.

Three other lineages, which are distributed in Turan and North Caucasus, may be provisionally placed in correspondence to the following combinations of subspecies: lineage IV – *D. s. lagopus*, *D. s. usuni*, *D. s. megacranius*; lineage Va – *D. s. innae*, *D. s. turanicus*, *D. s. austroruralensis*; lineage Vb – *D. s. nogai*. All these subspecies are attributed to the western (‘lagopus’) group. The above preliminary taxonomic assessment should be extended to include additional data on as yet insufficiently sampled subspecies. In particular, it remains to be confirmed whether any of the western subspecies are of relatively recent origin.

As a general conclusion, mitochondrial lineages are found to show relatively good correlation with morphological subspecies, and this correlation is more pronounced in the eastern than in the western part of the range. At the same time, the subdivision into the ‘sagitta’ and ‘lagopus’ subspecies groups is not supported. The Qaidam lineage apart, there are five highly divergent clades, the relationships between which are unresolved. Genetic distances between these lineages fall within the range typical for closely related species or subspecies (Baker & Bradley, 2006). The estimated age of radiation among the main clades (~500 kya–1 Myr) conforms to interspecific rather than intraspecific divergence. However, given that cytonuclear discordance is a common phenomenon (e.g. Ropiquet & Hassanin, 2006; Melo-Ferreira *et al.*, 2012), any taxonomic implications of this result should be treated with caution. Besides, as follows from the species tree, the available nuclear data,

albeit preliminary, are consistent with the ‘sagitta’–‘lagopus’ dichotomy. An extended sampling of genes and populations is needed to provide a comprehensive molecular taxonomic analysis of *Dipus*. Nevertheless, based on the present data, we conclude that *D. sagitta* is a species complex, which consists of two or more species.

ACKNOWLEDGEMENTS

We are grateful to Riccardo Castiglia and three anonymous reviewers for their helpful and constructive comments on the manuscript. We thank the Joint Russia-Mongolian Complex Biological Expedition of the Russian Academy of Sciences and the Academy of Sciences of Mongolia for support of the field work in Mongolia. We are grateful to Adiya Yadamsuren and Undrahbayar for their help with collecting jerboas. Many thanks to Prof. Kanat Akhmetov (Pavlodar State University, Kazakhstan) and Dr Bibigul Zhumabekova (Pavlodar State Pedagogical Institute) for collaboration during field work in Kazakhstan. This work was supported by the Russian Foundation for Basic Research, project 17-04-00065a (collection of material and genetic studies) and Russian Science Foundation, project 14-50-00029 (phylogenetic analysis and the processing of the paper). This is publication no. 938 of the Mitrani Department of Desert Ecology.

REFERENCES

- Baker RJ, Bradley RD. 2006.** Speciation in mammals and the genetic species concept. *Journal of Mammalogy* **87**: 643–662.
- Ben Faleh A, Granjon L, Tatard C, Boratyński Z, Said K, Cosson JF. 2012a.** Phylogeography of the Greater Egyptian jerboa *Jaculus orientalis* (Rodentia: Dipodidae) in Mediterranean North Africa. *Journal of Zoology* **286**: 208–220.
- Ben Faleh A, Granjon L, Tatard C, Ski ZB, Cosson JF, Said K. 2012b.** Phylogeography of two cryptic species of African desert jerboas (Dipodidae: *Jaculus*). *Biological Journal of the Linnean Society* **107**: 27–38.
- Carstens BC, Pelletier TA, Reid NM, Satler JD. 2013.** How to fail at species delimitation. *Molecular Ecology* **22**: 4369–4383.
- Cheng J, Lv X, Xia L, Ge D, Zhang Q, Lu L, Yang Q. 2017.** Impact of orogeny and environmental change on genetic divergence and demographic history of *Dipus sagitta* (Dipodoidea, Dipodinae) since the Pliocene in Inland East Asia. *Journal of Mammal Evolution*. doi: 10.1007/s10914-017-9397-6.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012.** Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**: 1969–1973.

- Dupanloup I, Schneider S, Excoffier L. 2002.** A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* **11**: 2571–2581.
- Fomin SV, Lobachev VS. 1988.** Confirmation of species autonomy and new finds of *D. halli* in Mongolia. Natural conditions and resources of some areas of the Mongolian People's Republic. Prirodnyye usloviya i resursy nekotorykh rayonov Mongol'skoy Narodnoy Respubliki Proceedings of the Conference Irkutsk: 114–115 (in Russian).
- Fu YX. 1997.** Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.
- Fujisawa T, Barraclough TG. 2013.** Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Systematic Biology* **62**: 707–724.
- Han W, Fang X, Berger A. 2012.** Tibet forcing of mid-Pleistocene synchronous enhancement of East Asian winter and summer monsoons revealed by Chinese loess record. *Quaternary Research* **78**: 174–184.
- Head MJ, Gibbard PL. 2015.** Early–Middle Pleistocene transitions: linking terrestrial and marine realms. *Quaternary International* **389**: 7–46.
- Heckel G, Burri R, Fink S, Desmet JF, Excoffier L. 2005.** Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. *Evolution* **59**: 2231–42.
- Heled J, Drummond AJ. 2010.** Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**: 570–580.
- Ho SYW, Shapiro B, Phillips MJ, Cooper A, Drummond AJ. 2007.** Evidence for time dependency of molecular rate estimates. *Systematic Biology* **56**: 515–522.
- Ito M, Jiang W, Sato JJ, Zhen Q, Jiao W, Goto K, Sato H, Ishiwata K, Oku Y, Chai JJ, Kamiya H. 2010.** Molecular phylogeny of the subfamily Gerbillinae (Muridae, Rodentia) with emphasis on species living in the Xinjiang-Uygur Autonomous Region of China and based on the mitochondrial cytochrome *b* and cytochrome *c* oxidase subunit II genes. *Zoological Science* **27**: 269–278.
- Jaarola M, Searle JB. 2002.** Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. *Molecular Ecology* **11**: 2613–2621.
- Jobb G. 2008.** *TREEFINDER, version of October 2008 [computer software and manual]*. Munich: Author. Available at: www.treefinder.de
- Lebedev VS, Bannikova AA, Pages M, Pisano J, Michaux JR, Shenbrot GI. 2012.** Molecular phylogeny and systematics of Dipodoidea: a test of morphology-based hypotheses. *Zoologica Scripta* **42**: 1–19.
- Li Q, Qiu ZD. 2005.** Restudies in Sminthoides Schlosser, a fossil genus of three-toed jerboa from China. *Vertebrata Palasiatica* **43**: 24–35.
- Liao J, Jing D, Luo G, Wang Y, Zhao L, Liu N. 2016.** Comparative phylogeography of *Meriones meridianus*, *Dipus sagitta*, and *Allactaga sibirica*: potential indicators of the impact of the Qinghai-Tibetan Plateau uplift. *Mammalian Biology* **81**: 31–39.
- Librado P, Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.
- Lv X, Xia L, Ge D, Wen Z, Qu Y, Lu L. 2016.** Continental refugium in the Mongolian Plateau during quaternary glacial oscillations: phylogeography and niche modelling of the endemic desert hamster, *Phodopus roborovskii*. *PLoS ONE* **11**: e0148182.
- Ndiaye A, Tatard C, Stanley W, Granjon L. 2016.** Taxonomic hypotheses regarding the genus *Gerbillus* (Rodentia, Muridae, Gerbillinae) based on molecular analyses of museum specimens. *Zookeys* **18**: 145–155.
- Neronov VM, Abramson NI, Warshavsky AA, Karimova TY, Khlyap LA. 2009.** Chorological structure of the range and genetic variation of the midday gerbil (*Meriones meridianus*). *Doklady Biological Sciences* **425**: 273–275.
- Melo-Ferreira J, Boursot P, Carneiro M, Esteves PJ, Farello L, Alves PC. 2012.** Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulations. *Systematic Biology* **61**: 367–381.
- Michaux JR, Libois R, Filippucci MG. 2005.** So close and so different: comparative phylogeography of two small mammal species, the yellow-necked fieldmouse (*Apodemus flavicollis*) and the woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. *Heredity* **94**: 52–63.
- McDevitt AD, Yannic G, Rambau RV, Hayden TJ, Searle JB. 2010.** Postglacial recolonization of Continental Europe by the pygmy shrew (*Sorex minutus*) inferred from mitochondrial and Y chromosomal DNA sequences. In: Habel JC, Assmann T, eds. *Relict species*. Berlin, Heidelberg: Springer, 217–236.
- Mohammadi S, Afonso S, Adibi MA, Melo-Ferreira J, Campos R. 2016.** A new and highly divergent mitochondrial lineage in the small five-toed Jerboa, *Allactaga elater*, from Iran (Mammalia: Rodentia). *Zoology in the Middle East* **62**(3).
- Montgelard C, Bentz S, Tirard C, Verneau O, Catzeflis FM. 2002.** Molecular systematics of Sciurognathi (Rodentia): the mitochondrial cytochrome *b* and 12S rRNA genes support the Anomaluroidea (Pedetidae and Anomaluridae). *Molecular Phylogenetics and Evolution* **22**: 220–233.
- Paradis E, Claude J, Strimmer K. 2004.** APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.
- Paupério J, Herman JS, Melo-Ferreira J, Jaarola M, Alves PC, Searle JB. 2012.** Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Eurasia. *Molecular Ecology* **21**: 6015–6032.
- Petrova TV, Tesakov AS, Kowalskaya YM, Abramson NI. 2016.** Cryptic speciation in the narrow-headed vole *Lasiopodomys (Stenocranius) gregalis* (Rodentia: Cricetidae). *Zoologica Scripta* **45**: 618–629.
- Petrova TV, Zakharov ES, Samiya R, Abramson NI. 2015.** Phylogeography of the narrow-headed vole *Lasiopodomys (Stenocranius) gregalis* (Cricetidae, Rodentia) inferred

- from mitochondrial cytochrome *b* sequences: an echo of Pleistocene prosperity. *Journal of Zoological Systematics and Evolutionary Research* **53**: 97–108.
- Pisano J, Condamine FL, Lebedev V, Bannikova A, Quéré J-P, Shenbrot GI, Pagès M, Michaux JR. 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.
- Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP. 2006.** Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* **55**: 595–609.
- Puillandre N, Lambert A, Brouillet S, Achaz G. 2012.** ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* **21**: 1864–1877.
- Qiu ZD, Wang XM, Li Q. 2006.** Faunal succession and biochronology of the Miocene through Pliocene in Nei Mongol (Inner Mongolia). *Vertebrata Palasiatica* **44**: 164–181
- Rambaut A, Drummond A. 2005.** *Tracer version 1.5. Computer program distributed by the authors.* Oxford: Department of Zoology, University of Oxford. Available at: <http://http://evolve.zoo.ox.ac.uk/software.html>
- Rogers AR, Harpending H. 1992.** Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**: 552–569.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012.** MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.
- Ropiquet A, Hassanin A. 2006.** Hybrid origin of the Pliocene ancestor of wild goats. *Molecular Phylogenetics and Evolution* **41**: 395–404.
- Sambrook J, Fritsch EF, Maniatis, T. 1989.** *Molecular cloning: a laboratory manual.* New York: Cold Spring Harbor Lab. Press.
- Shenbrot GI. 1991a.** Geographic variability of the horned jerboa *Dipus sagitta* (Rodentia, Didodidae). The general nature of intraspecific variability and subspecies differentiation in the western part of the species range. *Russian Journal of Zoology* **70**: 101–110 (in Russian).
- Shenbrot GI. 1991b.** Geographic variability of the horned jerboa *Dipus sagitta* (Rodentia, Didodidae). Subspecific differentiation in the territory of East Kazakhstan, Tuva and Mongolia. *Russian Journal of Zoology* **70**: 91–97 (in Russian).
- Shenbrot GI, Bannikova AA, Giraudoux P, Quéré J-P, Raoul F, Lebedev VS. 2017.** A new recent genus and species of three-toed jerboas (Rodentia: Dipodinae) from China: alive fossil? *Journal of Zoological Systematics and Evolutionary Research* **55**: 356–368.
- Shenbrot GI, Krasnov BR, Rogovin KA. 1999.** *Spatial ecology of desert rodent communities.* Berlin: Springer.
- Shenbrot GI, Sokolov VE, Geptner VG, Kovalskaya UM. 1995.** *Mammals of Russia and adjacent regions. Dipodidae.* Moscow: Nauka (in Russian).
- Stephens M, Donnelly P. 2003.** A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* **73**: 1162–1169.
- Stephens M, Smith NJ, Donnelly P. 2001.** A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**: 978–989.
- Swofford DL. 2003.** PAUP* – phylogenetic analysis using parsimony (*and other methods), Version 4. Sunderland: Sinauer Associates.
- Tajima F. 1989.** Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- Thomas O. 1908.** A new jerboa from China. *Annals and Magazine of Natural History* **2**: 307.
- Wakeley J. 1999.** Nonequilibrium migration in human history. *Genetics* **153**: 1863–1871.
- Yang Z. 2007.** PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.
- Yang DY, Eng B, Wayne JS, Dudar JC, Saunders SR. 1998.** Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *American Journal of Physical Anthropology* **105**: 539–543.
- Zazhigin VS, Lopatin AV. 2001.** The history of the Dipodoidea (Rodentia, Mammalia) in the Miocene of Asia: 4. Dipodinae at the Miocene-Pliocene transition. *Paleontological Journal* **35**: 60–74 (translated from *Paleontologicheskii Zhurnal* **1**: 61–75).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Geographical information, number and designation of specimens used in the study.

Table S2. Sequences of *Dipus sagitta* retrieved from GenBank.

Table S3. The original primers for amplification and sequencing of short fragments of *Cytb*.

Figure S1. Saturation plots for different substitution classes/codon positions of the *Cytb* gene. The X-axis represents the maximum likelihood distance estimated in PAUP* based on all substitution types at the 1st and 2nd codon positions and transversions at the 3rd codon positions. The Y-axis shows the uncorrected p-distance calculated from (A) separately transitions (ti) and transversions (tv) at the 3rd codon positions, and (B) all substitution types at the 1st and 2nd codon positions. Saturation is clearly present in 3rd codon position transitions. This follows from the fact that the slope of the linear regression fitted to the segment with ML distance >6% is close to zero.

Figure S2. The NJ tree of *Dipus sagitta* as inferred from the *Cytb* data on 228 specimens including all museum specimens (marked by arrows) and three species of *Stylodipus* as the outgroup.

Figure S3. Distribution of the uncorrected p-distances between *Cytb* haplotypes within *Dipus sagitta*.

Figure S4. The ML tree of the relationships among the alleles of *BRCA1* in *Dipus sagitta*. The colours correspond to the mitochondrial lineages in [Figure 2](#).

Figure S5. The ML tree of the relationships among the alleles of *IRBP* in *Dipus sagitta*. The colours correspond to the mitochondrial lineages in [Figure 2](#).