

Genetic evidence for several cryptic species within the *Scarturus elater* species complex (Rodentia: Dipodoidea): when cryptic species are really cryptic

ANNA BANNIKOVA^{1*}, VLADIMIR LEBEDEV², ANNA DUBROVSKAYA¹,
EVGENIA SOLOVYEVA², VIKTORIA MOSKALENKO², BORIS KRYŠTUFEK³,
RAINER HUTTERER⁴, ELENA BYKOVA⁵, BIBIGUL ZHUMABEKOVA⁶,
KONSTANTIN ROGOVIN⁷ and GEORGY SHENBROT⁸

¹Lomonosov Moscow State University, Vorobiev Gory, 119992 Moscow, Russia

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

³Slovenian Museum of Natural History, PO Box 290, 1001 Ljubljana, Slovenia

⁴Das Zoologische Forschungsmuseum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany

⁵Institute of Zoology Tashkent, Uzbekistan Academy of Science, 100125 Tashkent, Yunusabad district, Bogishamol str., 232 B, Uzbekistan

⁶Pavlodar State Pedagogical Institute, Mira str., 60 140002 Pavlodar, Kazakhstan

⁷Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Leninskii pr., 33, Moscow, Russia

⁸Mitrani Department of Desert Ecology, Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, 84990 Midreshet Ben-Gurion, Israel

Received 12 July 2018; revised 16 September 2018; accepted for publication 16 September 2018

Phylogeographical study of the small five-toed jerboa (*Scarturus elater*) and examination of the phylogenetic position of *S. vinogradovi* were performed using the mitochondrial cytochrome *b* (*cytb*) gene and fragments of the *BRCA1* and *IRBP* nuclear genes. Phylogenetic analysis of the *cytb* data including 115 specimens of *S. elater* from 47 localities across the species range revealed the existence of three highly divergent (10–11.3%) genetic clades: North (N), South (S) and South-West (SW). The N and S clades are well supported by nuclear genes and occur in sympatry across a large part of the range south of the Aral Sea. We found no trace of admixture between these clades, which suggests their reproductive isolation. We detected no morphological differences in the skull or glans penis between these two lineages, which we consider to represent an intriguing example of cryptic species. Given the reciprocal monophyly and deep genetic divergence, the SW lineage also deserves full species rank. The data indicate that *S. vinogradovi* is not a close relative of *S. elater*. It is placed as a separate deep branch in a clade also containing *S. elater s.l.* and *S. williamsi* + *S. euphratica*.

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

INTRODUCTION

The problem of cryptic species is currently the subject of intense debate because of its significance in understanding speciation processes and correctly assessing biodiversity (Bickford *et al.*, 2007; Jörger & Schrödl, 2013; Heethoff, 2018; Struck *et al.*, 2018a, b). Among

numerous definitions of the term ‘cryptic’ (‘hidden’) species, the most widely accepted is that ‘two or more species are “cryptic” if they are, or have been, classified as a single nominal species because they are at least superficially morphologically indistinguishable’ (Bickford *et al.*, 2007: p.149). Baker & Bradley (2006) defined cryptic species in mammals as those that would probably not be recognized based solely on classical studies of morphology of voucher specimens housed in museums.

*Corresponding author. E-mail: hylomys@mail.ru

Based on recent molecular studies, a number of species of small mammals with large geographical ranges have become recognized as complexes of ‘cryptic’ species (Neronov *et al.*, 2009; Ben Faleh *et al.*, 2012a, b; Paupério *et al.*, 2012; Petrova *et al.*, 2016; Lebedev *et al.*, 2018). Usually, genetic variation across the range of such species complexes is geographically structured, which requires only changes in taxonomic ranking, specifically a simple elevation of a subspecies level to a species. Such cases are generally not classified as truly cryptic species. Cases where morphologically similar, but genetically different species of rodents are sympatric are relatively rare. Moreover, in some cases detailed examination revealed that morphological characters allow for unambiguous identification of such species, which therefore are also not really cryptic (Shenbrot *et al.*, 2016). Most cases of true cryptic species among rodents are North American (Baker & Bradley, 2006). Among Palearctic desert rodents, we know of only one example of a true cryptic species, the *Gerbillus pyramidum* complex (Granjon *et al.*, 1999; Ndiaye *et al.*, 2016).

Jerboas of the *Scarturus elater* (Lichtenstein, 1825) species complex are the smallest (body mass 40–70 g) members of the five-toed jerboa subfamily Allactaginae. Traditionally, they were considered members of the genus *Allactaga* Cuvier, 1837. However, reconstruction of the molecular phylogeny of jerboas (Lebedev *et al.*, 2012) demonstrated polyphyly of *Allactaga* s.l., and the *elater* species complex was transferred to the genus *Paralactaga* Young, 1927. Recently it was suggested that *Paralactaga* should be treated as the junior synonym of *Scarturus* Gloger, 1841 and, thus, the *elater* species complex should be considered within the genus *Scarturus* (Michaux & Shenbrot, 2017). Considering the lack of a comprehensive revision of Allactaginae, this provisional decision is followed here.

The *S. elater* species complex is widely distributed across Palearctic deserts and semi-deserts from Cis- and Trans-Caucasia in European Russia, Armenia, Georgia and Turkey in the west (44°E) to Mongolian and Chinese Dzhungaria in the east (93°E), and from the lower parts of the Volga and Ural River valleys in the north (50°N) to the deserts of southern Iran, Afghanistan and Pakistan in the south (28°N). Across this extensive area, the species complex is well differentiated morphologically into three ‘morphotypes’ differing in molar size and morphology of the glans penis; these three morphotypes represent two species (*S. elater* with two groups of subspecies and *S. vinogradovi*). The first species is subdivided into two groups of subspecies (*elater* and *indica* groups) based on morphology of the glans penis. These groups are further subdivided into seven subspecies: four are in *elater*

group and three are in the *indica* group (Supporting Information, Fig. S1) based on variation in skull morphometry (Shenbrot, 1993; Shenbrot *et al.*, 1995). The recently described *Allactaga toussi* Darvish *et al.*, 2008 from northern Iran is similar to *S. vinogradovi* in the structure of male genitalia. Based on this feature, it was considered a subspecies of *S. vinogradovi* (Michaux & Shenbrot, 2017). At the same time, only within a relatively small part of its geographical range (northern Iran), the *S. elater* species complex is subdivided into four sympatric or parapatric genetic forms of species rank (Moshtaghi *et al.*, 2016). These data indicate that at least one species pair can be true cryptic species.

The aim of the present study was to examine the phylogeographical structure of the *S. elater* species complex across its entire geographical range with the focus on the correlation between recognized morphological species and subspecies and genetic variation to determine the existence of true cryptic species and to estimate the timing of divergence among the major geographical lineages.

MATERIAL AND METHODS

SAMPLING, DNA EXTRACTION AMPLIFICATION AND SEQUENCING

Most samples were obtained from fieldwork in Kazakhstan, Mongolia and Uzbekistan over the period 2012–2017 by the joint parasitology expedition of Pavlodar State Pedagogical Institute, Pavlodar University and Severtsov Institute of Ecology and Evolution RAN, Joint Russian–Mongolian Biological expedition and complex biological expedition of ZOOAZIA Company of Uzbekistan.

The entire sample consisted of 115 specimens of *S. elater* from 47 localities across the species range and one *S. vinogradovi* (Fig. 1, Table 1). In most cases, DNA was extracted from muscle tissue preserved in ethanol or small tissue biopsies of live-trapped animals. Seven specimens, including *S. vinogradovi* (ZMMU S-148087), the holotype of *S. e. strandi* Heptner, 1934 (ZMMU S-6729), a topotype of *S. e. heptneri* Pavlenko et Denisov, 1976 (ZMMU S-107490), *S. e. cf. elater* from southern Turkmenistan (S-137644), *S. e. cf. turkmeni* from south-west Turkmenistan (S-137638) and *S. e. cf. indica* from south-east Turkmenistan (S-145769) and Afghanistan (ZFMK 93.619) were represented by archive DNA extracted from museum skins or bones (age range 80–30 years). Overall, the original material included 96 specimens of *S. elater* and one specimen of *S. vinogradovi* and five specimens of other jerboas (*Allactaga major*, *A. severtsovi*, *Allactodipus bobrinskii*, *Pygeretmus pumilio* and *P. zhitzkovi*). Two

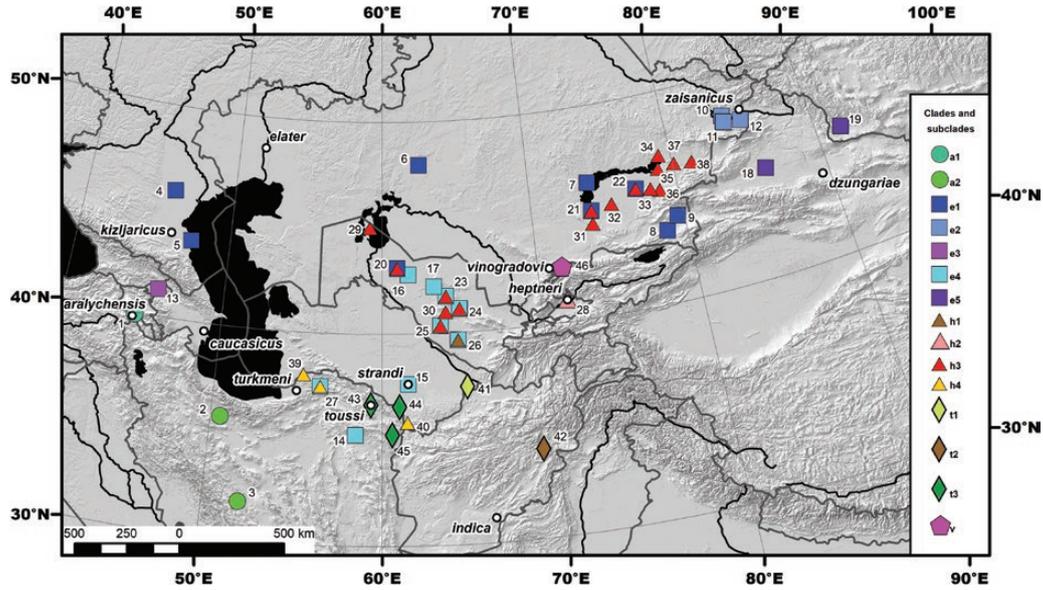


Figure 1. Geographical distribution of sampling localities and *cytb* lineages. Symbols are coloured by mitochondrial lineage as in Figure 2. Locality names and detailed geographical information are given in Table 1.

Table 1. Designation, tissue, geographic and genotype information of specimens used in the study. Genotypes of *BRCA1* and *IRBP* in which the allelic phase could not be determined with adequate posterior probability (<80%) are marked in italic.

locality (Fig.1)	Collecting locality (region or closest city and coordinates)	Subspecies (synonyms) following Shenbrot (1993) and mt lineage	Specimen code or/and <i>cytb</i> GB AcNo (Fig. 2–4)	Collection and museum code	Source	<i>Cytb</i> haplotype code	<i>BRCA1</i> genotype code	<i>IRBP</i> genotype code
1	Armenia (n=1) Probably, between Artashat and Vedi N 39° 56', E 44°38'	<i>aralychensis</i> ● a1	Armenia	ZMMU197177	orig. data: muscles	A1.1	br(16;16)	ir(20;20)
2	Iran 1 (n=3) Tehran, Iran N 35° 43.45', E 50° 33.91'	"turkmeni" ● a2	JQ954935 JQ954936 JQ954937	FUMZM2679 FUMZM2680 FUMZM2681	GeneBank	A2.2 A2.3 A2.4		
3	Iran 2 (n=1) Mirabad, Esfahan, Iran N 31° 48.38', E 52° 1.81'	"turkmeni" ● a2	JQ954934	FUMZM2678	GeneBank	A2.1		
4	Kalmykia (n=1) Chernozemelsk District, Kalmykia, Russia N 45°26'12", E 45°59' 51"	<i>elater</i> (<i>kizljarius</i>) ■ e1	Kalm2219	IEE RAN	orig. data: muscles	E1.3	br(25;25)	ir(21;33)

Table 1. (Continued)

locality (Fig.1)	Collecting locality (region or closest city and coordinates)	Subspecies (synonyms) following Shenbrot (1993) and mt lineage	Specimen code or/and <i>cytb</i> GB AcNo (Fig. 2–4)	Collection and museum code	Source	<i>Cytb</i> haplotype code	<i>BRCA1</i> genotype code	<i>IRBP</i> genotype code
5	Dagestan (n=1) Chakanny, Kizlyar district N 43° 47' 42", E 47° 31' 35"	<i>elater</i> ■e1 (<i>kizljarius</i>)	Dag30 KM397178	(uncatalogized, coll. L.Khlyap)	orig. data: muscles	E1.17	br(25;25)	ir(21;21)
6	Turgay (n=2) Lower Turgay River, Amangeldy, Aqtobe region, Kazakhstan N 48° 04' 18", E 62° 30' 16"	<i>elater</i> ■e1	K12-201 K12-266	Turgay01 Turgay66	orig. data: muscles muscles	E1.4 E1.5	br(25;20)	ir(23;26)
7	W Balkhash 1 (n=2) NW shore of Lake Balkhash, 30 km SW from Gulshad village, Qaraghandy region, Kazakhstan N 46° 23' 04.97", E 74° 00' 59.56"	<i>elater</i> ■e1	Kb16-26 Kb16-34	ZMMU197245 ZMMU197250	orig. data: muscles muscles	E1.6 E1.7	br(17;17) br(25;20)	ir(23;23) ir(13;24)
8	Ili River Hollow 1 (n=4) Syugatinskaya valley, 3 km south of the Kokpek village, Almaty region, Kazakhstan N 43°26'45.3", E 78°42'53"	<i>elater</i> ■e1	Kb16-71 Kb16-73 Kb16-75 Kb16-76	ZMMU197266 ZMMU197268 ZMMU197270 Kb16-76	orig. data: muscles muscles muscles ear biopsy	E1.12 E1.13 E1.14 E1.13	br(17;21) br(17;17) br(17;21) br(17;21)	ir(21;21) ir(21;21) <i>ir(21;21)</i> ir(21;21)
9	Ili River Hollow 2 (n=1) 4 km East of Aidarli village, Almaty region, Kazakhstan N 44° 02' 28.19", E 79° 34' 38.61"	<i>elater</i> ■e1	Kb16-84	ZMMU197271	orig. data: muscles	E1.15	Br(22;22)	ir(21;21)
10a	Zaisan 1 (n=2) N shore of Lake Zaisan, Shackelmes, East Kazakhstan region N 48° 04' 07", E 84° 08' 18"	<i>zaisanicus</i> ■e2	K12-66 K12-67	ZMMU190823 ZMMU190824	orig. data: muscles muscles	E5.1 E5.1	br(24;24)	ir(13;13)
10b	N 48° 04' 08", E 84° 08' 19" (n=1)	<i>zaisanicus</i> ■e2	K13-84	ZMMU192600	orig. data: muscles	E5.2	br(24;24)	ir(14;13)

Table 1. (Continued)

locality (Fig.1)	Collecting locality (region or closest city and coordinates)	Subspecies (synonyms) following Shenbrot (1993) and mt lineage	Specimen code or/and <i>cytb</i> GB AcNo (Fig. 2–4)	Collection and museum code	Source	<i>Cytb</i> haplotype code	<i>BRCA1</i> genotype code	<i>IRBP</i> genotype code
11a	Zaisan 2 (n=1) 7 km SE from the village Topolovy Cape, East Kazakhstan region, N 47° 47' 28", E 84° 07' 36"	<i>zaisanicus</i> ■e2	K12-74	ZMMU190825	orig. data: muscles	E5.3		
11b	Zaisan 3 (n=5) NW of Tugyl, East Kazakhstan region, N 47° 46' 17", E 84° 9' 58"	<i>zaisanicus</i> ■e2	K13-24 K13-25 K13-26 K13-27 K13-28	ZMMU192593 ZMMU192594 ZMMU192595 ZMMU192596 ZMMU192597	orig. data: muscles	E5.4 E5.4 E5.5 E5.4 E5.4	br(23;23) br(24;23)	ir(13;13)
12	Zaisan 4 (n=2) NE from Ulken-Karatal, East Kazakhstan region, N 47° 38' 54", E 85° 20' 57"	<i>zaisanicus</i> ■e2	K13-49 K13-50	ZMMU192598 ZMMU192599	orig. data: muscles muscles	E5.4 E5.4	br(24;23) br(23;23)	ir(13;13)
13	Georgia (n=3) Iori River valley, Dalis Mta Reservoir, Kakhet'i, Georgia N 41° 16' 56.10", E 45° 51' 01.14"	<i>caucasicus</i> ■e3	G17-1 G17-2 G17-3	G1 G2 G3	orig. data: muscles muscles muscles	E3.1 E3.2 E3.3	br(26;26) br(26;26) br(26;26)	ir(19;19) ir(18;19) ir(18;18)
14	Iran 3 (n=2) Kashmar, Khorasan-e Razavi, Iran N 35° 13.94', E 58° 27.48'	"turkmeni" or <i>indica</i> ■e4	JQ954932 JQ954931	FUMZM1429 FUMZM2128	GeneBank	E4.14 E4.15		
15	Turkmenistan 4 (n=1) Karabata, Mary	<i>elater (strandii)</i> ■e4	S-6729*	ZMMU 6729	orig. data: dried skin	E4.13		
16	Karakalpakstan 1 (n=2) Delta of the Ahchadarya dry river, The Republic of Karakalpakstan, Uzbekistan N 42° 51' 13.5", E 061° 21' 29.9"	<i>elater (strandii)</i> ■e4	Uz17-44 Uz17-55	ZMMU198807 ZMMU198808	orig. data: muscles muscles	E4.8 E4.9	br(29;32) br(29;33)	ir(26;31) ir(26;29)

Table 1. (Continued)

locality (Fig.1)	Collecting locality (region or closest city and coordinates)	Subspecies (synonyms) following Shenbrot (1993) and mt lineage	Specimen code or/and <i>cytb</i> GB AcNo (Fig. 2–4)	Collection and museum code	Source	<i>Cytb</i> haplotype code	<i>BRCA1</i> genotype code	<i>IRBP</i> genotype code
17	Navoi 3 (n=1) 30 km to the West from Uchkuduk, Navoi region, Uzbekistan N 42° 12' 33.4", E 063° 17' 03.4"	<i>elater (strandii)</i> ■e4	Uz17-23	ZMMU198801	orig. data: muscles	E4.6	br(29;32)	ir(26;27)
18	China (n=1) Musowan Field Station, 1.5 km NE, Xinjiang, China N 45° 07' 31.98", E 86° 01' 39.9"	<i>dzungariae</i> ■e5	Xin12 A1	Xin12 A1	orig. data: muscles	E5.1	br(27;27)	ir(17;17)
19	Khovd aimak (n=1) Bayan-Inder, 37 km SW from Wench, Khovd aimak, Mongolia N 47° 48' 36", E 91° 41' 42"	<i>dzungariae</i> ■e5	M15-131	ZMMU196374	orig. data: muscles	E5.2	br(27;27)	ir(16;16)
20	Karakalpakstan 2 (n=4) Delta of the Ahchadarya dry river, 15 km SE from Bozgul, The Republic of Karakalpakstan, Uzbekistan N 42° 59' 20.2", E 061° 19' 07.4"	<i>elater (strandii)</i> ■e1 ■e1 ■e4 ▲h3	Uz17-33 Uz17-34 Uz17-36 Uz17-37	ZMMU198802 ZMMU198803 ZMMU198805 ZMMU198806	orig. data: muscles muscles muscles muscles	E1.1 E1.2 E4.7 H3.1	br(29;29) <i>br(25;28)</i> <i>br(29;31)</i> br(3;3)	ir(26;32) ir(26;31) ir(29;31) ir(1;1)
21	W Balkhash 2 (n=4) SW shore of Lake Balkhash, Peninsula Burubaytal, Zhambyl Region, Kazakhstan N 45°02'44.9", E 74°00'37.6"	<i>elater</i> ■e1 ■e1 ■e1 ▲h3	Kb16-40 Kb16-41 Kb16-42 Kb16-43	ZMMU197251 ZMMU197252 ZMMU197253 ZMMU197254	orig. data: muscles muscles muscles muscles	E1.8 E1.9 E1.10 E1.11	<i>br(21;19)</i> <i>br(25;18)</i> br(9;1)	<i>ir(13;25)</i> <i>ir(13;25)</i> <i>ir(7;1)</i>

Table 1. (Continued)

locality (Fig.1)	Collecting locality (region or closest city and coordinates)	Subspecies (synonyms) following Shenbrot (1993) and mt lineage	Specimen code or/and <i>cytb</i> GB AcNo (Fig. 2–4)	Collection and museum code	Source	<i>Cytb</i> haplotype code	<i>BRCA1</i> genotype code	<i>IRBP</i> genotype code
22	S Balkhash 1 (n=2) left bank of the Karatal river, 10 km west of Naimansuek village, Almaty Region, Kazakhstan N 45°41'05.7", E 77°12'30.5"	<i>elater</i> ■e1 ▲h3	Kb16-108 Kb16-109	ZMMU197272 ZMMU197273	orig. data: muscles muscles	E1.16 H3.18	br(17;21)	ir(21;15)
23	Navoi 1 (n=2) Sands Zhamankum, 30 km North-Northwest from Zarafshan, Navoi region, Uzbekistan N 41° 44' 15.1", E 064° 01' 27.2"	<i>elater (strandii)</i> ▲h3 ■e4	Uz17-18 Uz17-19	ZMMU198799 ZMMU198800	orig. data: muscles muscles	H3.1 E4.5	br(13;1) br(29;29)	ir(8;2) ir(26;27)
24	Navoi 2 (n=5) Karakatinskaya hollow, Navoi region, Uzbekistan N 41° 07' 07.6", E 064° 50' 21.6"	<i>elater (strandii)</i> ■e4 ▲h3 ■e4 ■e4 ■e4	Uz17-05 Uz17-06 Uz17-07 Uz17-08 Uz17-09	ZMMU198794 ZMMU198795 ZMMU198796 ZMMU198797 ZMMU198798	orig. data: muscles muscles muscles muscles muscles	E4.1 E4.2 E4.3 E4.4 E4.4	br(1;2)	ir(1;1) ir(26;26) ir(26;30) ir(26;30)
25	Bukhara 1 (n=3) Takyrs to S from Kuldzhuktau, Bukhara region, Uzbekistan N 40° 32' 27.3", E 063° 46' 56.2"	<i>elater (strandii)</i> ■e4 ■e4 ▲h3	Uz17-71 Uz17-74 Uz17-75	ZMMU198811 ZMMU198812 ZMMU198813	orig. data: muscles muscles muscles	E4.10 E4.3 H3.4	br(29;32) br(2;2)	ir(26;28) ir(1;5)
26	Bukhara 2 (n=4) Bukhara Hazard cattery, Bukhara region, Uzbekistan N 39° 38' 10", E 064° 39' 3.6"	<i>elater (strandii)</i> ■e4 ■e4 ▲h1 ▲h1	Uz17-82 Uz17-83 Uz17-80 Uz17-81	ZMMU198817 ZMMU198818 ZMMU198815 ZMMU198816	orig. data: muscles muscles muscles muscles	E4.11 E4.12 H1.1 H1.3	br(29;33)	ir(26;26) ir(10;10) ir(1;15) ir(1;10)
27	Iran 4 (n=4) Golestan NP, Khorasan-e Shemali, Iran N 37° 28.106', E 56° 18.048'	<i>turkmeni</i> ▲h4 ▲h4 ▲h4 ■e4	JQ954927 JQ954929 JQ954930 JQ954928	FUMZM2675 FUMZM2676 FUMZM2677 FUMZM2674	GeneBank	H4.1 H4.2 H4.3 E4.16		

Table 1. (Continued)

locality (Fig.1)	Collecting locality (region or closest city and coordinates)	Subspecies (synonyms) following Shenbrot (1993) and mt lineage	Specimen code or/and <i>cytb</i> GB AcNo (Fig. 2–4)	Collection and museum code	Source	<i>Cytb</i> haplotype code	<i>BRCA1</i> genotype code	<i>IRBP</i> genotype code
28	Fergana valley (n=1) Namangan, Uzbekistan	<i>elater (heptneri)</i> ▲h2	S-107490*	ZMMU107490	orig. data: dried skin	H2.1		
29	Karakalpakstan 3 (n=2) Vozrojdeniye Island, Uzbekistan N 45° 01' 00", E 59° 10' 00"	<i>elater</i> ▲h3	Uz17-1479 Uz17-1480		orig. data: muscles muscles	H3.5 H3.5	br(2;2) br(10;1)	ir(9;1)
30	Bukhara 3 (n=2) S slope of Kuldzhuktau, Bukhara region, Uzbekistan N 40° 45' 02.0", E 063° 49' 46.2"	<i>elater (strandii)</i> ▲h3	Uz17-66 Uz17-68	ZMMU198809 ZMMU198810	orig. data: muscles muscles	H3.2 H3.3	br(2;2) br(1;4)	ir(1;6) <i>ir(1;11)</i>
31	Chu-Ili Mountains (n=13) NE slope of the Chu-Ili Mountains, 30 km NE of Hantau, Zhambyl Region, Kazakstan N 44°26'12.8", E 73°58'48.6"	<i>elater</i> ▲h3	Kb16-44 Kb16-45 Kb16-46 Kb16-49 Kb16-50 Kb16-51 Kb16-53 Kb16-52 Kb16-54 Kb16-55 Kb16-60 Kb16-61 Kb16-62	Kb16-44 Kb16-45 Kb16-46 ZMMU197255 ZMMU197256 ZMMU197257 ZMMU197259 ZMMU197258 ZMMU197260 ZMMU197261 ZMMU197262 ZMMU197263 ZMMU197264	orig. data: ear biopsy ear biopsy ear biopsy muscles muscles muscles muscles muscles muscles muscles muscles muscles	H3.6 H3.7 H3.8 H3.9 H3.10 H3.11 H3.6 H3.12 H3.13 H3.14 H3.15 H3.16 H3.17	br(1;3)	ir(1;4) <i>ir(1;12)</i>
32	S Balkhash 2 (n=2) Bozbor, Almaty re- gion, Kazakstan N 45° 15' 26", E 75° 25' 31.6"	<i>elater</i> ▲h3	K15-07 K15-08	ZMMU196178 ZMMU196179	orig. data: muscles muscles	H3.29 H3.30	br(1;1)	ir(2;12)
33	S Balkhash 3 (n=1) N of Mulala, Almaty region, Kazakhstan N 45° 32' 39", E 78° 11' 24"	<i>elater</i> ▲h3	K15-22	ZMMU196182	orig. data: muscles	H3.31	br(1;1)	

Table 1. (Continued)

locality (Fig.1)	Collecting locality (region or closest city and coordinates)	Subspecies (synonyms) following Shenbrot (1993) and mt lineage	Specimen code or/and <i>cytb</i> GB AcNo (Fig. 2–4)	Collection and museum code	Source	<i>Cytb</i> haplotype code	<i>BRCA1</i> genotype code	<i>IRBP</i> genotype code
34	E Balkhash 1 (n=2) 40 km NW from Aktogay village, Almaty Region, Kazakhstan N 46° 59' 51", E 79° 10' 20"	<i>elater</i> ▲h3	Kb16-135 Kb16-136	ZMMU197280 ZMMU197281	orig. data: muscles muscles	H3.25 H3.24	br(1;1) br(1;2)	ir(1;3) ir(1;1)
35	E Balkhash 2 (n=1) SE shore of Lake Balkhash to the north of the Sarykurak, Almaty region, Kazakhstan N 46° 23' 02.8", E 78° 57' 23.7"	<i>elater</i> ▲h3	K15-27	ZMMU196183	orig. data: muscles	H3.32	br(1;1)	
36a	S Balkhash 4 (n=2) 4 km NE from the Kyzylagash vil- lage, Almaty Region, Kazakhstan N 45°24'39.2", E 78°46'36.3"	<i>elater</i> ▲h3	Kb16-118 Kb16-121	ZMMU197274 Kb16-121	orig. data: muscles ear biopsy	H3.19 H3.20	br(1;3)	<i>ir(1;1)</i>
36b	E Balkhash 3 (n=2) Akbalyk, SW from Aktogay village, Almaty region N 46°33'00", E 79°13'53"	<i>elater</i> ▲h3	AR K16-01 AR K16-02	ZMMU196804 ZMMU196805	orig. data: muscles muscles	H3.26 H3.27		
37a	E Balkhash 4 (n=1) Mountains of Arkaly, the hill of Zhosaly, Almaty Region, Kazakhstan N 46°29'02.6", E 80°06'19"	<i>elater</i> ▲h3	Kb16-130	ZMMU197279	orig. data: muscles	H3.24	br(2;3)	ir(1;3)
37b	E Balkhash 5 (n=1) Sarykum, SE from Aktogay village, Almaty region N 46°40'32", E 80°13'24"	<i>elater</i> ▲h3	AR K16-06	ZMMU196806	orig. data: muscles	H3.28	br(11;1)	

Table 1. (Continued)

locality (Fig.1)	Collecting locality (region or closest city and coordinates)	Subspecies (synonyms) following Shenbrot (1993) and mt lineage	Specimen code or/and <i>cytb</i> GB AcNo (Fig. 2–4)	Collection and museum code	Source	<i>Cytb</i> haplotype code	<i>BRCA1</i> genotype code	<i>IRBP</i> genotype code
38	Alakul' Hollow (n=3) SE shore of Koshkarkol Lake, Almaty Region, Kazakhstan N 46° 21' 36.95", E 81° 19' 13.92"	<i>elater</i> ▲h3	Kb16-125 Kb16-126 Kb16-127	ZMMU197277 ZMMU197278 Kb16-127	orig. data: muscles muscles ear biopsy	H3.21 H3.22 H3.23	br(3;7) br(1;1) br(2;8)	ir(1;1) ir(1;3) ir(1;1)
39	Turkmenistan 3 (n=1) Krasnovodsk Region, Sumbar, Kysil-Atrek	<i>turkmeni</i> ▲h4	S-137638	ZMMU137638	orig. data: dried skin	H4.5		
40	Turkmenistan 2 (n=1) Ashgabad reg., Serakh, AkarChashme	<i>elater</i> ▲h4	S-137644	ZMMU137644	orig. data: dried skin	H4.4		
41	Turkmenistan 1 (n=1) Obruchev Step, Ish-Kak, 10 km E, Lebap, N 37.4146, E 65.0711	<i>indica</i> (?) ◆t1	S-145769	ZMMU145769	orig. data: dried skin	T2.1		
42	Afghanistan (n=1) Dasht above Lagar-Tal, 35 km S of Kabul, Afghanistan N 34.2215, E 69.1469	<i>indica</i> ◆t2	BK 16	ZFMK 93.619	orig. data: bone	T1.1		
43	Mashhad (n=6) Cheshme Gilas, Khorasan-e Razavi, Iran N 36° 38.1', E 59° 20.1'	<i>toussi</i> ◆t3	JQ954957 JQ954938 JQ954958 JQ954954 JQ954955 JQ954956	FUMZM1415 FUMZM1416 FUMZM1418 FUMZM2694 FUMZM2695 FUMZM2696	GeneBank	T3.1 T3.2 T3.3 T3.4 T3.5 T3.6		
44	Sarakhs (n=2) Sarakhs, Khorasan-e Razavi, Iran N 36° 30.35', E 61° 7.02'	<i>toussi</i> ◆t3	JQ954933 JQ954959	FUMZM1431 FUMZM2130	GeneBank	T3.7 T3.8		

sequences of exon 11 of the breast cancer type 1 susceptibility protein (*BRCA1*), four sequences of the interphotoreceptor retinoid-binding photoreceptor (*IRBP*) and 26 sequences of the mitochondrial

locality (Fig. 1)	Collecting locality (region or closest city and coordinates)	Subspecies (synonyms) following Shenbrot (1993) and mt lineage	Specimen code or/and <i>cytb</i> GB AcNo (Fig. 2–4)	Collection and museum code	Source	<i>Cytb</i> haplotype code	<i>BRCA1</i> genotype code	<i>IRBP</i> genotype code
45	Turbat Jam (n=1) Turbat Jam, Khorasan-e Razavi, Iran N 35° 13.971', E 60° 35.742'	<i>toussi</i> ◆t3	AJ389534	T-1045	GeneBank	T3.9		
46	Kyrgyzstan (n=1) Kirovskoye, Talas district, Kyrgyzstan N 42,6430, E 71,5821	<i>Scarturus</i> <i>vinogradovi</i>	S-148087*	ZMMU148087	orig. data: dried skin and bone			
47	Unknown loc. (n=1)	<i>elater</i> ■e1	Zoo	-	orig. data: muscles	E1.18	br(25;25)	
48	Unknown loc. (n=1)	<i>elater</i>	Zoo2013	ZMMU194246	orig. data: muscles	H0		
	Turkey (n=1) Konya, N37°51' E32°29'	<i>Scarturus</i> <i>williamsi</i>	KC465439	Kon1	GeneBank			
	Turkey (n=1) Harran, N36°51' E39°01'	<i>Scarturus</i> <i>euphratica</i>	KC465442 (1)	Har1	GeneBank			
	Syria (n=1) Homs, Karyatien, N34°07' E37°13'	<i>Scarturus</i> <i>aulacotis</i>	KC465446 (2)	Kar1	GeneBank			
	Iran (n=1) Esfahan, Peykan	(?) <i>Scarturus</i> <i>hotsoni</i>	JQ954943	FUMZM2685	GeneBank			
	E Balkhash 5 (n=1) Sarykum, SE from Aktogay village, Almaty region N 46°40'32", E 80°13'24"	<i>Pygeretmus</i> <i>pumilio</i>	AR K16-05	ZMMU196834	orig. data: muscles			
	Alakul' Hollow (n=1) SE shore of Koshkarkol Lake, Almaty Region, Kazakhstan N 46° 21' 36.95", E 81° 19' 13.92"	<i>Pygeretmus</i> <i>zhitkovi</i>	Kb16-124	ZMMU197356	orig. data: muscles			

locality (Fig.1)	Collecting locality (region or closest city and coordinates)	Subspecies (synonyms) following Shenbrot (1993) and mt lineage	Specimen code or/and <i>cytb</i> GB AcNo (Fig. 2–4)	Collection and museum code	Source	<i>Cytb</i> haplotype code	<i>BRCA1</i> genotype code	<i>IRBP</i> genotype code
W Balkhash 1 (n=1)	NW shore of Lake Balkhash, 30 km SW from Gulshad village, Qaraghandy Region, Kazakhstan N 46° 23' 04.97", E 74° 00' 59.56"	<i>Allactaga major</i>	Kb16-23	ZMMU197282	orig. data: muscles			
S Balkhash 1 (n=1)	left bank of the Karatal river, 10 km west of Naimansuek village, Almaty Region, Kazakhstan N 45°41'05.7", E 77°12'30.5"	<i>Allactaga severtsovi</i>	Kb16-111	ZMMU197286	orig. data: muscles			
	Kyzylkum desert, Uzbekistan	<i>Allactaga severtsovi</i>	AVA2006	(uncatalogized, coll. Abramov)	orig. data: muscles			
Navoi 3 (n=1)	30 km to the West from Uchkuduk, Navoi region, Uzbekistan N 42° 12' 33.4", E 063° 17' 03.4"	<i>Allactodipus bobrinskii</i>	Uz17-22	ZMMU198829	orig. data: muscles			
Mongolia, Khovd aymag (n=1)		<i>Orientallactaga sibirica</i>	164 MF076854	ZMMU181016	orig. data: muscles			
Mongolia, Dundgovi aymag (n=1)		<i>Orientallactaga bullata</i>	141 KM397179	ZMMU179572	orig. data: muscles			
Mongolia, Bayankhongor aymag, Ekhiingol (n=1)	N 43°14'45", E 98°59'58"	<i>Orientallactaga balikunica</i>	M2005-125 KM397180	M2005_125	orig. data: muscles			

*type or topotype specimens

cytochrome *b* (*cytb*) gene (including 19 sequences of *S. elater*) were retrieved from GenBank (Supporting Information, Table S1).

Total DNA from ethanol-preserved tissues from liver, kidney and ear clippings was extracted using a

standard protocol of proteinase K digestion, phenol-chloroform deproteinization and isopropanol precipitation (Sambrook *et al.*, 1989).

For most specimens, we sequenced the complete *cytb* gene and fragments of two nuclear loci: *BRCA1*

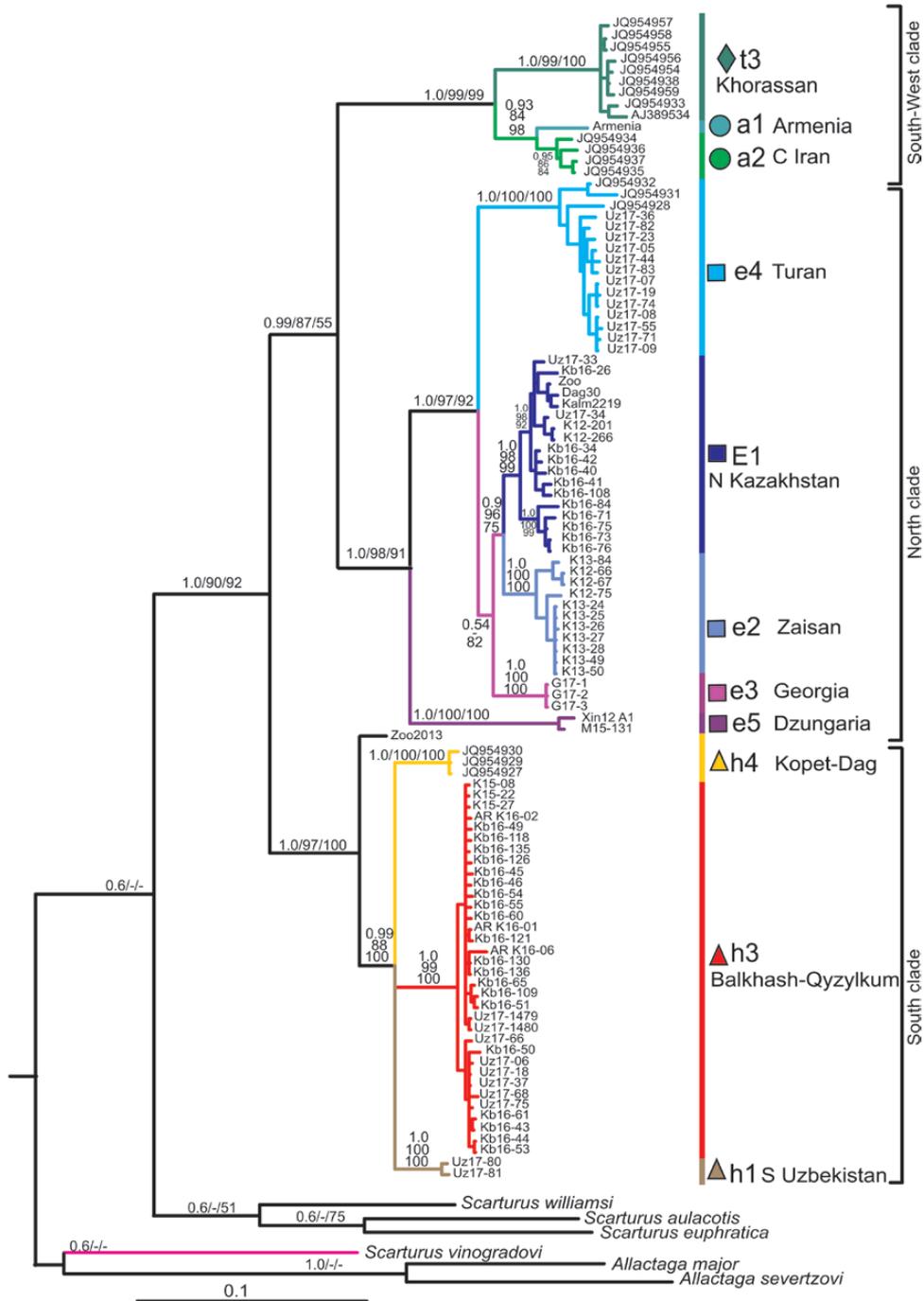


Figure 2. Bayesian phylogeny of the *S. elater* species complex as inferred from *cytb* data of 110 specimens, including four other species of *Scarturus* and two species of *Allactaga* as the outgroup. Numbers above/below branches correspond to MrBayes posterior probabilities, ML and MP bootstrap support (> 50%) for the main clades. Scale bar denotes 0.1 substitutions per site.

and *IRBP*. External primers and polymerase chain reaction (PCR) protocols are described in [Lebedev et al. \(2012, 2018\)](#) and [Pisano et al. \(2015\)](#). For sequencing the *cytb* gene of *S. elater*, a combination of internal

primers H601ae/L400ae specifically designed for this study ([Table S2](#)) was used. For sequencing the *IRBP* gene, we used R1175 and R701dip primers from [Lebedev et al. \(2012\)](#); for sequencing *BRCA1*,

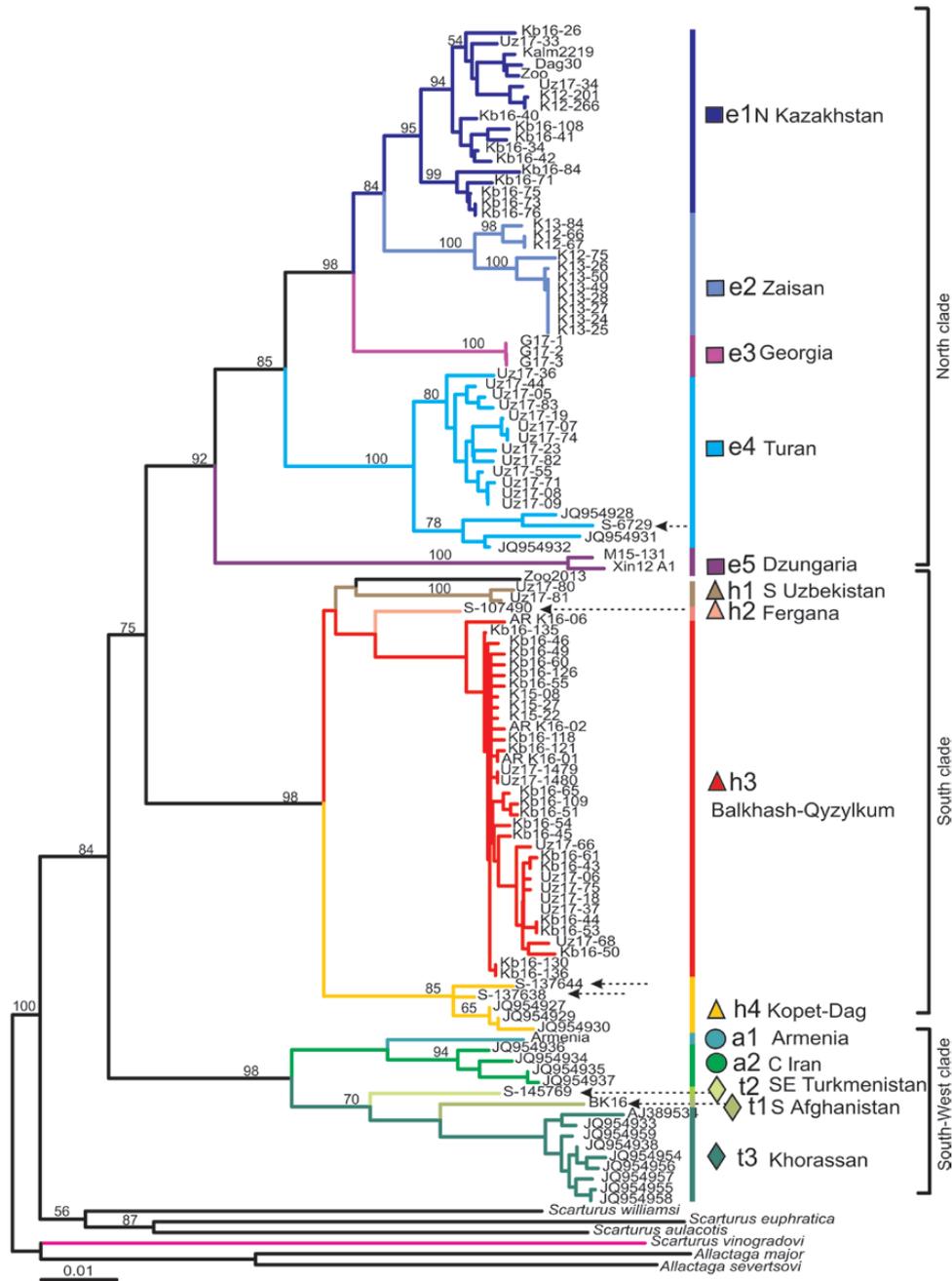


Figure 3. NJ tree of the *S. elater* species complex as inferred from *cytb* sequences of 116 specimens, including shorter fragments (<300 bp) obtained from six museum specimens (marked by arrows). Numbers above branches denote bootstrap support (1000 pseudoreplicates). Scale bar denotes 0.01 substitutions per site.

the external primers from [Lebedev *et al.* \(2012, 2018\)](#) were used.

DNA from dried skins and bones of museum collections were purified directly using the MiniElute PCR Purification Kit (Qiagen), including an overnight lysis step following the manufacturer’s protocol and the

recommendations of [Yang *et al.* \(1998\)](#). DNA extracted from museum specimens was highly degraded; thus, for most specimens, only short five or six overlapped fragments (80–260 bp) of *cytb* were obtained using the H601_Ae and L400_Ae internal primers and combinations of additional primers specifically designed for

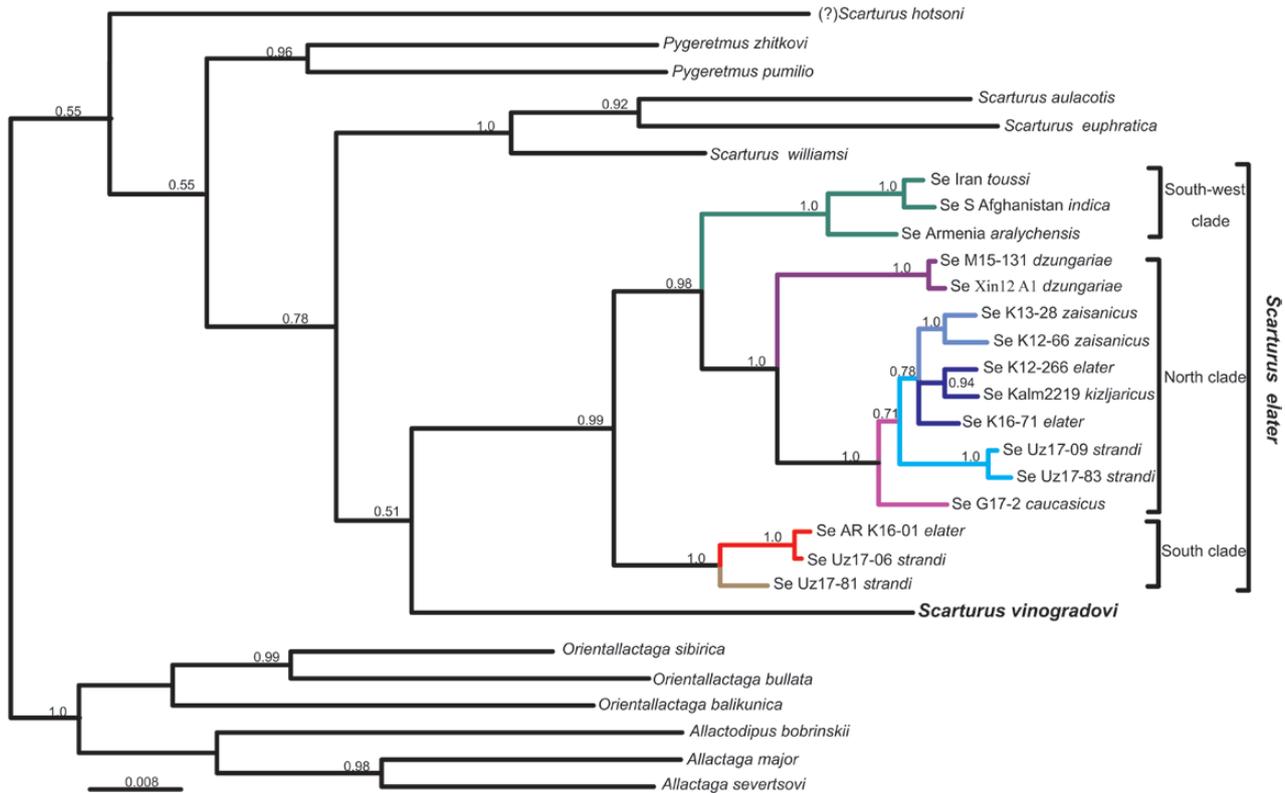


Figure 4. Bayesian phylogeny of the genus *Scarturus* as inferred from all substitutions at the 1st and 2nd codon positions and transversions at the 3rd codon position of the *cytb* alignment. The examined samples include selected representatives of the main clades of *S. elater* s.l. and species of *Allactaga*, *Allactodipus*, *Pygeretmus* and *Orientallactaga* as the outgroup. Scale bar denotes 0.008 substitutions per site.

this study (Table S2). For *S. e. indica* ZFMK 93.619, two non-overlapping *cytb* fragments (84 and 184 bp) were amplified in combinations of L400_Ae/H601_Ae and L543_Av/H751_Av. For *S. vinogradovi*, six overlapping *cytb* fragments of 80–200 bp and one *IRBP* fragment of 250 bp were obtained. The PCR programme for amplification of short fragments included an initial denaturation at 95 °C for 3 min, 45 cycles at 95 °C for 30 s, 52–55 °C (depending on primer) for 30 s and 72 °C for 30 s, and a final extension at 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 conducted 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.

The sequences obtained in this study can be accessed via GenBank (Accession numbers: *cytb* MH973322–MH973415, *IRBP* MH979160–MH979217, *BRCA1* MH991639–MH991673).

MITOCHONDRIAL DATA: TREE RECONSTRUCTIONS AND DISTANCE ESTIMATION

In the phylogenetic analysis of *S. elater* all substitutions for all codon positions were used. Trees were reconstructed under maximum likelihood (ML), maximum parsimony (MP) and Bayesian information (BI) criteria using sequences longer than 800 bp. The outgroup consisted of *S. williamsi*, *S. euphratica*, *S. aulacotis*, *S. vinogradovi*, *Allactaga major* and *A. severtsovi*. A second analysis [neighbour joining (NJ) tree reconstruction] was conducted using using a larger alignment including short *cytb* fragments (<300 bp) obtained for museum specimens.

To determine the phylogenetic position of *S. vinogradovi* we also performed an additional Bayesian analysis focusing on phylogenetic relationships among the main branches of Allactaginae and employing an extended set of taxa including *Scarturus hotsoni*, *Orientallactaga sibirica*, *O. bullata*, *O. balikunica*, *Allactodipus bobrinskii*, *Pygeretmus pumilio* and *P. zhitkovi*. In this case the transitions at the 3rd codon positions were excluded because of saturation.

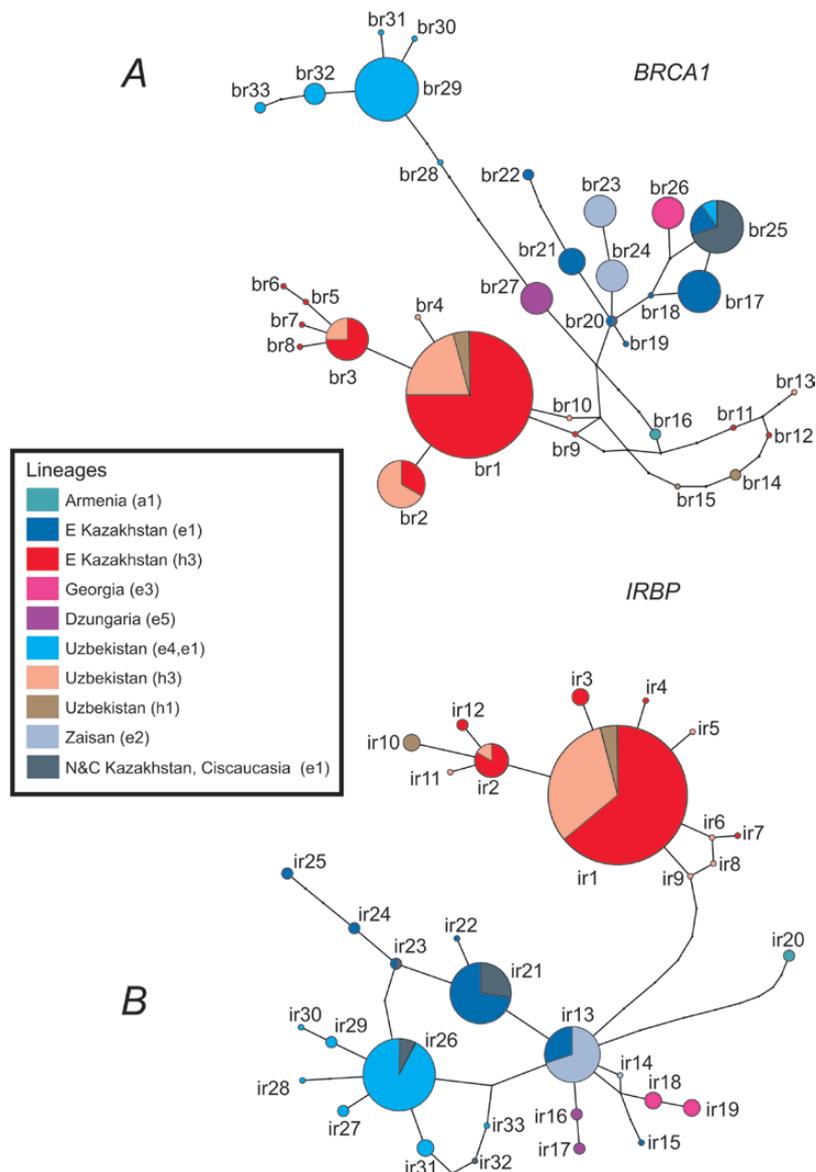


Figure 5. TCS network showing the relationships among the alleles of *BRCA1* (A) and *IRBP* (B) genes in *S. elater* s.l. Allele and specimen codes are given in Table 1. The size of circles corresponds to the number of specimens with identical alleles. Colours denote corresponding mitochondrial lineages (see Figs 2, 3).

ML reconstructions were conducted in IQTree v.1.6 (Nguyen *et al.*, 2015). The ModelFinder routine (Kalyaanamoorthy *et al.*, 2017) was used to determine the optimum partitioning scheme and best-fit substitution models for each subset under the Bayesian information criterion (BIC). Clade stability was tested using Ultrafast Bootstrap (Minh *et al.*, 2013) with 10 000 replicates.

A Bayesian tree reconstruction was conducted in MrBayes 3.2 (Ronquist *et al.*, 2012) assuming separate models for each of the codon positions. The analysis included two independent runs. Chain length was set

at 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.

Unweighted parsimony analysis was performed in PAUP* 4.0b10 (Swofford, 2003) invoking random addition sequence with 50 replicates and TBR branch swapping. Clade stability was estimated based on 1000 bootstrap replicates with ‘Multrees’ option not in effect. PAUP* 4.0b10 was also used to calculate the

Table 2. Approximate clade age estimates (Mya) in *S. elater* and outgroups based on mitochondrial data

Node of species, clades or subclades	Age (Mya)	95% HPD
<i>S. williamsi</i> + <i>S. euphratica</i> / <i>S. elater</i>	2.80	2.11–3.78
tmrca <i>S. elater</i> s.l.	1.26	0.96–1.65
tmrca N clade Dzungarian (e5)/((N Kazakh (e1) + Zaisan (e2) + Georgian (e3))/Turanian (e4))	0.56	0.41–0.73
Turanian (e4)/(N Kazakh (e1) + Zaisan (e2) + Georgian (e3))	0.25	0.19–0.32
tmrca (N Kazakh (e1) + Zaisan (e2) + Georgian (e3))	0.15	0.10–0.19
tmrca S clade	0.19	0.14–0.25
tmrca SW clade	0.28	0.20–0.38

HPD, highest posterior density; tmrca, time to the most recent common ancestor.

Table 3. Morphometric variability (SD) along the first three PCA axes of the subsamples from the areas of co-occurrence of the two clades and the subsamples from the areas where only one of the clades was recorded

Locality	Line	<i>N</i>	SD _{PC1}	SD _{PC2}	SD _{PC3}
NW Kazakhstan	e1	31	0.6169	0.9526	0.9940
Betpakdala and N Balkhash	e1	63	0.7630	0.8499	0.9491
SE Balkhash	h3	35	0.8595	0.7471	1.0262
Kopet-Dag	e4 + h4	72	0.8301	0.9129	0.9995
NW Kyzylkum	e1 + e4 + h3	36	0.6866	0.9899	0.8400
C and S Kyzylkum	e4 + h1 + h3	81	0.7257	0.7820	0.9246
S Balkhash	e1 + h3	82	0.7875	0.7343	0.9439

matrix of genetic distances among haplotypes (uncorrected p-distance) and to reconstruct the NJ tree.

NUCLEAR DATA

For allelic phase reconstruction, the Phase module (Stephens *et al.*, 2001; Stephens & Donnelly, 2003) implemented in the software DNAsp v.5 (Librado & Rozas, 2009) was used. Networks of haplotypes were reconstructed using TCS under default options (Clement *et al.*, 2000) and visualized using tcsBU (Múrias dos Santos *et al.*, 2015).

MOLECULAR DATING

The ultrametric tree was generated in BEAST v.1.84 (Drummond *et al.*, 2012) under both a strict and a relaxed log-normal clock. Partitioning and substitution models were set as in the ML analysis, the birth–death tree prior was employed. Based on the AICm statistic (a posterior simulation-based analog of Akaike's information criterion through Markov chain Monte Carlo: Baele *et al.*, 2012) the strict clock was preferred. Considering time-dependency of the molecular rate estimates, node heights were corrected using the method suggested by Bannikova *et al.* (2010). For this,

an additional analysis using only the third position transversions was performed. Estimates of the absolute ages were calculated based on the assumption that the basal radiation within the crown Allactaginae (i.e. the split between *Allactaga* and *Scarturus*) occurred in the latest Miocene (MN13), as supported by both the fossil record (Zazhigin & Lopatin, 2000) and molecular dating results (Shenbrot *et al.*, 2017). An exponential distribution with hard lower (5.3 Mya) and 95% soft upper (7.5 Mya) bounds was used as the prior density for the tree root age.

MORPHOLOGICAL DATA

Morphology of the glans penis was examined in 19 genotyped male specimens (Line e4: S-197245, S-197252, S-197266, S-197270, S-198794, S-198796, S-198801, S-198803, S-198807, S-198811, S-198812; Line h1: S-198816; Line h3: S-197256, S-197262, S-197273, S-197274, S-197279, S-198799, S-198809). Under 20-fold magnification, we measured the length, width and height of each glans, scored the pattern of spine distribution, and counted the number of spines. For analysis of morphometric skull variability, we used the data for 13 cranial measurements of 708 specimens of *S. elater*, grouped into 23 subsamples. Details of measurements and sample composition were provided

by Shenbrot (1993, 2009). The combined sample was analysed by principal components analysis (PCA) with normalized varimax rotation using STATISTICA for Windows release 7.0 package (StatSoft Inc., Tulsa, OK, USA). For each subsample with $N > 30$, the standard deviation was estimated along each PCA axis.

RESULTS

CHARACTERISTICS OF *CYTB* SEQUENCE

Among the 115 sequences of *S. elater*, we found 96 unique *cytb* haplotypes. For most specimens, the length of sequences was complete or near complete (890–1140 bp). For museum vouchers of *S. elater*, the length of five sequences was 175 bp and one sequence (S-137638154) was 154 bp. For specimen *S. e. indica* BK16, the total length of two fragments was 267 bp, and the total length of the *cytb* sequence of a museum specimen of *A. vinogradovi* was 843 bp. The number of nucleotide substitutions observed among sequences of *S. elater* were as follows: 3rd position, 748 substitutions; 1st, 107; and 2nd, eight.

THE *CYTB* PHYLOGENETIC TREE

We performed BI, ML and MP phylogenetic analyses of the complete and near complete *cytb* sequences (Fig. 2) and the NJ analysis of the extended sample including six short sequences of museum specimens (Fig. 3). In both cases, three highly supported clades were revealed. The level of divergence between these clades was 9.7–11.3% (p-distance). All of the clades were further structured into sublineages (phylogroups), differing by 3–9%, which mainly revealed strong geographical associations. Additionally, haplotypes belonging to two main clades were found in sympatry over a large part of their range (Fig. 1).

According to the BI tree (Fig. 2), two clades, named here South-West (SW) and North (N), were in a sister position. The SW clade included populations from the south-western part of the species range (Fig. 1) and was subdivided into the following sublineages: Armenian (a1), Central Iranian (a2) and Khorassan (t3) (Fig. 2). In the NJ tree, the SW clade also contained haplotypes from south-eastern Turkmenistan (t1) and southern Afghanistan (t2) (Fig. 3). The distance between t3 and a1 + a2 subclades was 6.8%.

The N clade included specimens from Uzbekistan, Kalmykia, East Kazakhstan, Georgia and the Xinjiang Province of China. It was subdivided into five sublineages differing by 3–9%: Northern Kazakhstan (e1), Zaisan (e2), Georgian (e3), Turanian (e4) and Dzungarian (e5). The last two lineages differed from the others by 5.8% and 8.6%, respectively.

The third major clade of *S. elater* in the BI tree [the South (S) clade] occupied Uzbekistan, south-eastern Kazakhstan and Turkmenistan and was further subdivided into four sublineages (pairwise p-distances of ~4%): South Uzbekistan (h1), Fergana (h2), wide-range Balkhash–Qyzylkum (h3) and Kopet-Dag (h4), and a separate branch formed by one specimen from an unknown locality (Zoo2013). The phylogroups of the S clade were allopatric; however, in Uzbekistan and southern Balkhash this lineage coexisted with the N clade (Fig. 1). Localities of sympatric occurrence of the N and S lineages are as follows: Navoi 1, Karakalpakstan 2, Bukhara 1, Bukhara 2 and southern Balkhash 1 (Table 1 and Fig. 1). Based on genetic differences between these three clades and moderate to no support for the branching pattern in ML, MP and NJ analyses, we interpret the tree topology as indicating a potential trichotomy.

Molecular data rejected the hypothesis that *S. vinogradovi* is close to the *S. elater* complex. Comparison with the available data for Allactaginae demonstrated that it is highly divergent from all known lineages of five-toed jerboas, including *Orientalallactaga*, *Allactodipus*, *Pygeretmus*, and (?) *Scarturus hotsoni* (Fig. 4). In the phylogenetic trees, it was either placed in a separate deep branch, which was more distant from *S. elater* than from *S. williamsi* + *S. euphratica* (Figs 2, 3), or emerged from the trichotomy with the latter two branches (Fig. 4). However, support for its phylogenetic position was low.

NUCLEAR DATA

BRCA1

Based on a sample of 65 sequences (767 bp) corresponding to 37 genotypes, 33 haplotypes (alleles) were reconstructed in PHASE. Allelic phase could not be determined with adequate posterior probability (> 0.8) in five sequences (Table 1). Observed heterozygosity was 0.54. Relationships among haplotypes as inferred by TCS are shown in Figure 5A (see also the NJ tree of unphased genotypes, Supporting Information, Fig. S2).

IRBP

Based on a sample of 57 sequences (1068 bp), 80 haplotypes in 36 genotypes were reconstructed in PHASE. Observed heterozygosity was 0.56. Allelic phase could not be determined with adequate posterior probability (> 0.8) in seven sequences (Table 1). The relationships among haplotypes are shown in Figure 5B. The sequence of *S. vinogradovi* was substantially divergent from all sequences of *S. elater* (Fig. S3).

The topology of *BRCA1* and *IRBP* networks summarizing the nuclear variability did not depart from

that retrieved by mitochondrial trees. The SW clade was represented in the nuclear DNA analysis by a single sample, but the remaining *cytb* clades (N and S) were characterized by additional internal structuring to the subclades. The substructuring was more pronounced in the *IRBP* tree, in which the N and S clades were separated by seven nucleotide substitutions, whereas eight substitutions distinguished the single sample belonging to the SW clade. Most importantly, no heterozygotes contained alleles from the N and S clades in the extensive zone of their sympatry.

MOLECULAR DATES

According to the results of the molecular clock analysis in BEAST (Table 2), the *S. elater* species complex diverged from other *Scarturus* taxa in the Late Pliocene (~2.8 Mya) while radiation of *S. elater s.l.* dates back to the second half of the Early Pleistocene (~1.3 Mya). Radiation was presumably rapid, with almost simultaneous separation of the three lineages. Within the N clade, the Dzungarian sublineage diverged in the Middle Pleistocene (~560 kya), while the Turanian branch split from other populations of *S. elater* (N Kazakhstan, Zaisan and Georgia) about 250 kya. The split of the SW clade into two branches (Armenia + C Iran) vs. Khorassan occurred in the Middle Pleistocene (~280 kya).

MORPHOLOGICAL DATA

We analysed the morphology of the glans penis of 11 specimens of the N clade and nine specimens of the S clade. Differences in measurements of the glans penis (mean \pm SD) differed between the two clades were statistically non-significant [length 3.23 ± 0.22 and 3.19 ± 0.18 mm ($t = 0.48$, $P = 0.64$), width 2.03 ± 0.23 and 2.22 ± 0.21 mm ($t = 1.85$, $P = 0.08$), height 1.14 ± 0.16 and 1.24 ± 0.12 mm ($t = 1.40$, $P = 0.18$)]. In specimens of both clades, the entire surface of the glans penis was uniformly covered by spines; the size of spines more or less evenly increased from top to the base. The total number of spines (mean \pm SD) was 23.27 ± 7.75 in clade N and 25.50 ± 10.53 in clade S (difference was statistically non-significant: $t = 1.07$, $P = 0.30$).

Results of PCA of morphometric data demonstrated that the first three principal components explained more than 65% of the total variation (Table S3). Comparison between sympatric and allopatric subsamples demonstrated comparable levels of morphometric variability (Table 3).

We also analysed the morphology of the glans penis of *S. vinogradovi* (in five specimens from Kirovskoye, Talas valley, Kirgizstan; detailed description given

in Shenbrot (1993) and in Shenbrot (2008)] and *S. toussi* (from a photograph provided by J. Darvish from the type locality – Cheshme Gilas, Khorasan Razavi Province, Iran, $36^{\circ}38'N$, $59^{\circ}19'E$) and could not detect any differences in size, location or number of spines. The glans penis of both specimens, *toussi* and *vinogradovi*, were cordate in plan, with pointed apex and well-expressed dorsal longitudinal furrow bifurcating to the tip. The glans penis length was 4.4–4.7 mm in *S. vinogradovi* and 4.2 mm in *S. toussi*; width was 3.0–3.1 and 2.9 mm, respectively. In both taxa, the spines were arranged in oblique rows, directed from longitudinal furrows forward and outward; the size of spines decreased evenly from top to base and there were seven rows, each with two to four spines on the dorsal surface. The number of spines on the dorsal surface of the glans penis was 32–36 in *S. vinogradovi* and 34 in *S. toussi*.

DISCUSSION

GENERAL REMARKS

Our results revealed the existence of three highly divergent genetic lineages in the *S. elater* complex; two of them were sufficiently sampled and well supported by nuclear genes. In spite of partial sympatry of these phylogroups, we did not observe any signature of introgression. Moreover, the main finding of our study was the sympatric distribution of the S and N clades across a large part of their ranges. Sympatric populations of the N and S clades showed no admixture, and no specimens heterozygous for alleles of the two lineages were found. This suggests a lack of intergroup gene flow and confirms a deep genetic divergence and reproductive isolation between the N and S branches of *S. elater*. These entities evidently represent cryptic species rather than intraspecific lineages of mitochondrial DNA (mtDNA). Based on the reciprocal monophyly and deep genetic divergence among the three clades, the SW lineage deserves full species rank. Moreover, considering that each clade is deeply structured, the true number of cryptic species may be even larger. Only a few sublineages of the three main clades were known earlier (Dianat *et al.*, 2013; Moshtaghi *et al.*, 2016), from north and central Iran. None of the previous studies included a geographically representative sample encompassing the entire range of the species. Our work includes a thorough geographical sampling of *S. elater* that allowed us to clarify the relationship between known genetic lineages, to establish their taxonomic status using not only mtDNA but also nuclear loci, and to propose a hypothesis for their evolutionary history.

Another important finding is the isolated position of *S. vinogradovi*, which appeared to be unrelated to *S. elater* s.l. The mitochondrial results (Fig. 4) suggest that it may form a clade together with *S. elater* and *S. williamsi*; accurate inference of its phylogenetic position would be possible based on a multigene analysis.

PHYLOGEOGRAPHICAL STRUCTURE IN THE *S. ELATER* COMPLEX AND ITS TAXONOMIC IMPLICATIONS

The valid taxonomic name for the N clade is *S. elater* Lichtenstein, 1825, which was described from West Kazakhstan. This lineage is characterized by the most pronounced internal differentiation. Among five subclades found within it, only the Turanian (e4) sublineage was previously known (=‘*A. elater* 1’ sensu Moshtagi *et al.*, 2016). The Turanian and Dzungarian (e5) lineages represent the deepest branching of the N clade tree. According to the *cytb* tree, the Dzungarian subclade is more distant from other lineages than the Turanian subclade, being the most basal branch among the whole N clade. In contrast, the analysis of nuclear loci implies that the Turanian subclade is the most divergent. Given the *cytb* genetic distance and nuclear gene divergence, separate species ranks for both Turanian and Dzungarian subclades may be suggested. The clear cytonuclear discordance is observed between the distribution of mitochondrial and nuclear haplotypes of N Kazakh (e1) and Turanian (e4) sublineages in Karakalpakia (Uzbekistan, loc20 in Fig. 1). At this locality mitotypes of both e1 and e4 lineages were found (e1: Uz17-33, Uz17-34; e4: Uz17-36). Uz17-36 has both mitochondrial and nuclear haplotypes of the Turanian sublineage, while Uz17-33 and Uz17-34 possess North Kazakh mtDNA but Turanian nuclear alleles (Figs 2, 3, 5, Table 1) indicating gene flow between these sublineages.

The possible taxonomic names for the subtaxa within the N clade are as follows: North Kazakh (e1) sublineage – *S. elater* s.s.; Zaisan (e2) sublineage – *zaisanicus*; Georgian (e3) sublineage – *caucasicus*; Turanian (e4) sublineage – *strandii* (as follows from the sequence of the holotype); Dzungarian (e5) sublineage – *dzungariae*.

The name of the species corresponding to the S clade is unclear; it is either *S. turkmeni* (Goodwin, 1940) described from Mazandaran (potentially sublineage h4) or *S. heptneri* (Pavlenko et Denisenko, 1976) described from Fergana Valley (sublineage h2 as revealed in a toptype). The problem is whether the type specimen of *turkmeni* belongs to the S clade or not. In the latter case (*turkmeni* can be a junior synonym of *strandii*), the only name available for this group is *S. heptneri*. The structure of the S clade is less

pronounced than that of the N clade. Additionally, the nuclear genotypes are known only for the Balkhash–Qyzylkum (h3) and southern Uzbekistan (h1) mitochondrial sublineages and were not obtained for the Kopet-Dag sublineage (h4 = ‘*A. elater* 2’ of Moshtagi *et al.*, 2016). Thus, until additional nuclear data on the expanded geographical sample are available, we suggest the S clade should be considered a single species.

The problem of naming the SW clade is that it is unclear whether the type of *indica* belongs to it, as might be expected from the fact that our sample was taken not far (approximately 400 km) from *terra typica* of the latter taxon. If this hypothesis is correct then the name of SW clade will be *S. indica* (Gray, 1842) as it is the senior synonym. Then, if the mtDNA phylogeny is correct, *S. toussi* should be treated as a junior synonym of *S. indica*. Five mitochondrial sublineages within the SW clade are combined into western (*aralychensis*) and eastern (*indica* + *toussi*) subgroups, the genetic distance between them possibly corresponding to interspecies divergence. Therefore, one may suggest the existence of an additional cryptic species, *S. aralychensis* (Satunin, 1901). However, in our analysis of nuclear genes, the SW clade was represented only by the Armenian specimen and only short fragments of museum archive DNA were obtained for lineages t1 and t2. Therefore, it is difficult to interpret with confidence the real taxonomic status of these subclades until additional data are obtained.

MORPHOLOGICAL VARIATION IN COMPARISON WITH MOLECULAR RESULTS

Earlier morphological analyses found differences between *S. elater* and *S. vinogradovi* in molar size and morphology of the glans penis, and between two groups of *S. elater* subspecies (‘*elater*’ and ‘*indica*’), which differed in morphology of the glans penis (Shenbrot, 1993, Shenbrot *et al.*, 1995). The recently described species *toussi* (Darvish *et al.*, 2008) is morphologically similar to *S. vinogradovi*. This prompted Michaux & Shenbrot (2017) to reclassify them as distinct subspecies of *S. vinogradovi*, with distribution ranges more than 1500 km apart. To explain the disparity between the morphological similarity of the two taxa and their deep genetic divergence, we propose three hypotheses. (1) Our estimates of morphological similarity and genetic dissimilarity are entirely correct, indicating an intriguing case of a pair of true cryptic species. (2) One may hypothesize that *S. toussi* is actually *S. vinogradovi*, which harbours mtDNA introgressed from the SW clade of *S. elater* s.l. In the absence of nuclear DNA data for *toussi*, it is difficult to conclude which of the two assumptions is true, although the second one appears less plausible. (3) Our current assessment

of morphological similarity of *S. toussi* and *S. vinogradovi* and dissimilarity of *S. toussi* and the SW clade of *S. elater* may be an overestimate. Additional data on skull morphometry of *S. toussi* published by Dianat *et al.* (2010) demonstrated that *S. toussi* is much more similar to *aralychensis* + *indica* than to *S. vinogradovi* (Table S4). The conclusion regarding the morphological similarity of *S. toussi* and *S. vinogradovi* is based mainly on the structure of glans penis. However, data on the penis morphology of *S. toussi* are based on only one photograph. Moreover, penile morphology in the *indica* group was studied in 15 specimens from south-east Turkmenistan, one specimen from central Iran and two specimens from Armenia (Shenbrot, 1993). For a more confident conclusion, it is necessary to examine additional material including specimens from type localities of *S. indica* and *S. toussi*. Thus, the question of the species rank of *S. toussi* is open, and hence the SW lineage possibly contains at least two separate species (*S. indica* and *S. aralychensis*). Concerning the 'elater' group of subspecies (*sensu* Shenbrot, 1993), morphological data have demonstrated the presence of four subspecies based on variation in skull morphometry but did not reveal any subdivision into two groups corresponding to the N and S genetic clades. Our data demonstrated that the morphology of the the glans penis does not really differentiate between these two clades. We did not have enough genetically identified specimens of the two genetic clades for statistical morphometric comparison. Nevertheless, if these two genetic clades differ in skull morphometry, samples containing a mixture of the two clades should be morphologically more variable than samples containing only one genetic clade. However, our comparison demonstrated that the level of morphometric variability of the samples from the areas of co-occurrence of the two clades was the same as that of the samples from the areas where only one of the clades was recorded. Therefore, these two clades within the *S. elater* species complex evidently represent true cryptic species.

Seven morphological subspecies are known for *S. elater* (Shenbrot, 1993), and some of these are geographically congruent with mitochondrial lineages. On the other hand, there are cases where morphologically identical populations belong to different mitochondrial lineages. Moreover, ranges of some subspecies overlap with several mitochondrial lineages. For example, the subspecies *caucasicus* and *zaisanicus* possess their own geographically localized mitochondrial haplotypes (respectively e3 and e2). Animals from north-western and central Iran, earlier assigned to *turkmeni* (Shenbrot, 1993), possess a separate *cytb* haplotype (SW: a2), thus representing a separate taxon that has to be formally described and named.

In the case of subspecies *dzungariae*, the western (SE Balkhash and Alakul' Hollow) and the eastern (Dzhungarian Basin) populations belong to different mitochondrial sublineages (h3 and e5, respectively). Populations of the Fergana Valley and Kyzylkum are morphologically similar but belong to different mitochondrial sublineages (h2 and h3 + e4, respectively). Within the range of *strandii* and *turkmeni*, mitochondrial sublineages h1, h3 and h4 of the S clade and e4 sublineage of the N clade are found.

POSSIBLE HISTORY OF THE RANGE

As follows from the molecular and morphological data, the relationships between the subspecies and genetic lineages are quite a challenge because of the complex geographical distribution of the genetic lineages. In each the localities from south-east Kazakhstan (except localities 21 and 22), only one form (N or S) is usually found. The S clade dominates in the south and east of the Balkhash region, whereas in the eastern and northern parts, only the N clade was found. The distribution of the N lineage is apparently disjunct as it occurs in the south-easternmost margin of the range in the middle Ili Valley (localities 8 and 9) where the S lineage is evidently absent. Therefore, range formation of the *S. elater* species group has a complex history.

Our molecular time estimations indicate that the split among the N, S and SW clades of *S. elater* is ancient, dating back to the second half of the Early Pleistocene (Calabrian). It may be hypothesized that the N and S lineages putatively responded to the climatic shifts of the Pleistocene in different ways. It is possible that during glacial phases, populations of the N clade dominated across the whole of the Balkhash region, whereas during periods of climate warming, the range of the N clade shifted northwards leaving behind refugional populations such as that in the Syugaty Valley (locality 8).

It is tempting to speculate that, in contrast to the N clade, populations of the S clade could have expanded in the area during interglacial periods and were fragmented during cold glacial phases of the Pleistocene. Different sublineages of the S clade may correspond to different glacial refugia. The wide range of the Balkhash–Qyzylkum lineage is probably a result of postglacial expansion from one such refugium. It is noteworthy that, in contrast to the Balkhash region, in Qyzylkum, the S lineage is widely sympatric with *strandii*, which is one of the main sublineages of the N clade. It remains unclear whether this sympatry was facilitated by niche segregation or, more plausibly, should be explained by a recent colonization of Qyzylkum by the S lineage (presumably during the Holocene).

The geographical distribution of the genetic lineages of the N clade indicates that its centre of origin could be in East Kazakhstan, whereas the north-western populations were formed during a recent wave of colonization. Another early colonization event was penetration of *S. elater* into Transcaucasia.

As already mentioned, the *Gerbillus pyramidum* complex represents the only documented case of a true cryptic species among Palearctic desert rodents (Granjon *et al.*, 1999; Ndiaye *et al.*, 2016). Our study on the *S. elater* species complex demonstrates at least two further examples of a deep genetic subdivision without morphological divergence in a widely distributed Palearctic desert rodent. We included a thorough geographical sampling of *S. elater* s.l. that allowed us to clarify the relationships between genetic lineages, to establish their taxonomic status, and to demonstrate that the true cryptic species are a real evolutionary phenomenon rather than gaps in our knowledge of morphological variation.

ACKNOWLEDGMENTS

We thank Prof. Kanat Akhmetov (Pavlodar State University, Kazakhstan) for collaboration during fieldwork in Kazakhstan, Evgeniy Peregontsev and Alexandr Schos (Zoocomplex LTD, Uzbekistan) for their help during fieldwork in Uzbekistan, and Alexey Surov for joint work during expeditions in Kazakhstan and Mongolia. Many thanks to 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 also thank three anonymous reviewers for their helpful comments. This work was supported by the Russian Foundation for Basic Research, projects 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).

REFERENCES

- Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. 2012. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution* **30**: 239–243.
- Baker RJ, Bradley RD. 2006. Speciation in mammals and the genetic species concept. *Journal of Mammalogy* **87**: 643–662.
- Bannikova AA, Lebedev VS, Lissovsky AA, Matrosova V, Abramson NI, Obolenskaya EV, Tesakov AS. 2010. Molecular phylogeny and evolution of the Asian lineage of vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial *cytb* sequence. *Biological Journal of the Linnean Society* **99**: 595–613.
- Ben Faleh A, Granjon L, Tatard C, Boratyński Z, Said K, Cosson JF. 2012b. 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. 2012a. Phylogeography of two cryptic species of African desert jerboas (Dipodidae: *Jaculus*). *Biological Journal of the Linnean Society* **107**: 27–38.
- Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* **22**: 148–155.
- Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**: 1657–1659.
- Darvish J, Matin MM, Haddad F. 2008. New species of five-toed jerboa (Rodentia: Dipodidae, Allactaginae) from north-east Iran. *Journal of Sciences, Islamic Republic of Iran* **19**: 102–109.
- Dianat M, Aliabadian M, Darvish J, Akbarirad S. 2013. Molecular phylogeny of the Iranian Plateau five-toed jerboa, *Allactaga* (Dipodidae: Rodentia), inferred from mtDNA. *Mammalia* **77**: 95–103.
- Dianat M, Tarahomi M, Darvish J, Aliabadian M. 2010. Phylogenetic analysis of the five-toed jerboa (Rodentia) from the Iranian Plateau based on mtDNA and morphometric data. *Iranian Journal of Animal Biosystematics* **6**: 49–59.
- 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.
- Granjon L, Bonnet A, Hamdine W, Volobouev V. 1999. Reevaluation of the taxonomic status of North African gerbils usually referred to as *Gerbillus pyramidum* (Gerbillinae, Rodentia): chromosomal and biometrical data. *Zeitschrift für Säugetierkunde* **64**: 298–307.
- Heethoff M. 2018. Cryptic species—conceptual or terminological chaos? A response to Struck *et al.* *Trends in Ecology & Evolution* **33**: 310–312.
- Jörger K, Schrödl M. 2013. How to describe a cryptic species? Practical challenges of molecular taxonomy. *Frontiers in Zoology* **10**: 1–27.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587–589.
- Lebedev VS, Bannikova AA, Lu L, Snytnikov EA, Yansanjav A, Solovieva EA, Abramov AV, Surov AV, Shenbrot GI. 2018. Phylogeographic study reveals high genetic diversity in a widespread desert rodent *Dipus sagitta* Pallas, 1773 (Dipodidae, Rodentia). *Biological Journal of the Linnean Society* **123**: 445–462.
- 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.
- Librado P, Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.
- Michaux J, Shenbrot G. 2017.** Family Dipodidae (Jerboas). In: Wilson DE, Lacher TE, Russell Jr, Mittermeier A, eds. *Handbook of the mammals of the world, 7. Rodents II*. Barcelona: Lynx Edicions in association with Conservation International and IUCN, **7**: 1–1008.
- Minh BQ, Nguyen MAT, von Haeseler A. 2013.** Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* **30**: 1188–1195.
- Moshtaghi S, Darvish J, Mirshamsi O, Mahmoudi A. 2016.** Cryptic species diversity in the genus *Allactaga* (Rodentia: Dipodidae) at the edge of its distribution range. *Folia Zoologica* **65**: 142–147.
- Múrias dos Santos A, Cabezas MP, Tavares AI, Xavier R, Branco M. 2015.** tcsBU: a tool to extend TCS network layout and visualization. *Bioinformatics* **32**: 627–628.
- Ndiaye A, Chevret P, Dobigny G, Granjon L. 2016.** Evolutionary systematics and biogeography of the arid habitat-adapted rodent genus *Gerbillus* (Rodentia, Muridae): a mostly Plio-Pleistocene African history. *Journal of Zoological Systematics and Evolutionary Research* **54**: 299–317.
- 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.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015.** IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution* **32**: 268–274.
- 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.
- 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.
- 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>.
- 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.
- Sambrook J, Green MR. 1989.** *Molecular cloning: a laboratory manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Shenbrot GI. 1993.** Revision of subspecies systematics of jerboas of the genus *Allactaga* of the USSR fauna. *Proceedings of the Zoological Institute, USSR Academy of Sciences* **243**: 81–109.
- Shenbrot GI. 2008.** *Mammals of Russia and adjacent regions*. Boca Raton: CRC Press.
- Shenbrot G. 2009.** On the conspecificity of *Allactaga hotsoni* Thomas, 1920 and *Allactaga firouzi* Womochel, 1978 (Rodentia: Dipodoidea). *Mammalia* **73**: 231–237.
- 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 G, Feldstein T, Meiri S. 2016.** Are cryptic species of the Lesser Egyptian Jerboa, *Jaculus jaculus* (Rodentia, Dipodidae), really cryptic? Re-evaluation of their taxonomic status with new data from Israel and Sinai. *Journal of Zoological Systematics and Evolutionary Research* **54**: 148–159.
- Shenbrot GI, Sokolov VE, Heptner VG, Koval'skaya YUM. 1995.** *Mammals of the fauna of Russia and contiguous countries. Dipodoid rodents*. Moscow: Nauka Press [in Russian]; English translation (2008). *Jerboas. Mammals of Russia and adjacent regions*, Hoffmann RS, Wilson DE, eds. Enfield: Science Publishers.
- Stephens M, Donnelly P. 2003.** Comparison of Bayesian methods for haplotype reconstruction from population genotype data. *The American Journal of Human Genetics* **73**: 1162–1169.
- Stephens M, Smith NJ, Donnell P. 2001.** A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics* **68**: 978–989.
- Struck TH, Feder JL, Bendiksbj M, Birkeland S, Cerca J, Gusarov VI, Kistenich S, Larsson K-H, Liow LH, Nowak MD, Stedje B, Bachmann L, Dimitrov D. 2018a.** Finding evolutionary processes hidden in cryptic species with molecular markers. *Trends in Ecology & Evolution* **33**: 153–163.
- Struck TH, Feder JL, Bendiksbj M, Birkeland S, Cerca J, Gusarov VI, Kistenich S, Larsson K-H, Liow LH, Nowak MD, Stedje B, Bachmann L, Dimitrov D. 2018b.** Cryptic species – more than terminological chaos: a reply to Heathoff. *Trends in Ecology & Evolution* **33**: 310–312.
- Swofford DL. 2003.** *PAUP* - Phylogenetic analysis using parsimony (*and other methods), version 4*. Sunderland: Sinauer Associates.
- 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. 2000.** The history of the Dipodoidea (Rodentia, Mammalia) in the Miocene of Asia: 3. Allactaginae. *Paleontological Journal* **34**: 553–565 [translated from *Paleontologicheskii Zhurnal* **5**: 82–94].

SUPPORTING INFORMATION

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

Table S1. Sequences of *Scarturus elater* and outgroup retrieved from GeneBank.

Table S2. The original primers for amplification and sequencing of short fragments of *cytb* and *IRBP* genes from the museum specimens of *S. elater* and *S. vinogradovi*.

Table S3. Mean principal components scores for eight samples of *Scarturus elater s.l.* Phylogenetic lineages E1 and H3 are allopatric while E + H denotes their sympatric occurrence. Reported also are factor loadings, eigenvalues for each principal component (PC) and the proportion of variance explained by each component.

Table S4. Comparison of cranial measurements of *Scarturus elater*, *S. vinogradovi* and *S. toussi*.

Figure S1. The putative range of subspecies of the *S. elater* species complex according to [Shenbrot \(1993\)](#): four subspecies in the *elater* group (A) and three subspecies in the *indica* group (B).

Figure S2. NJ tree illustrating the relationships among the unphased genotypes of *BRCA1* in *S. elater*. Colours correspond to the mitochondrial lineages in [Figures 2](#) and [3](#).

Figure S3. NJ tree illustrating the relationships among the unphased genotypes of *IRBP* in *S. elater*. Colours correspond to the mitochondrial lineages in [Figures 2](#) and [3](#).