Russian J. Theriol. 23 (1): 18–24

Complete mitochondrial genomes of Asian endemic white-toothed shrews: *Crocidura armenica* and *C. serezkyensis* (Eulipotyphla: Soricidae)

Leonid L. Voyta*, Tatyana V. Petrova, Valentina A. Panitsina, Semyon Yu. Bodrov & Natalia I. Abramson

ABSTRACT. The subfamily Crocidurinae unites white-toothed shrews (Eulipotyphla: Soricidae) and is the largest group of living eulipotyphlans; among them, the genus *Crocidura* contains approximately half of soricids diversity. Within the *Crocidura* endemics of Central and West Asia, species group "pergrisea" retains relentless interest. In this context, a major question is the taxonomic position and validity of an Armenia endemic *Crocidura armenica*, which since 2014 has been excluded from the list of Mammal Species of the World. In this study, we obtained the complete mitochondrial genome of *C. armenica* from a holotype. In addition, we determined partial mitochondrial genomes of a Pamir Mountains endemic *C. serezkyensis* and of a white-toothed shrew with unclear taxonomic status: *Crocidura* cf. *pergrisea* from Nakhchivan, Azerbaijan. The phylogenetic analysis with the inclusion of the mitogenome sequences of the 28 species showed the closest position of *C. armenica* and *Crocidura* cf. *pergrisea*, with a sister position of the *C. serezkyensis* specimen.

How to cite this article: Voyta L.L., Petrova T.V., Panitsina V.A., Bodrov S.Yu., Abramson N.I. 2024. Complete mitochondrial genomes of Asian endemic white-toothed shrews: *Crocidura armenica* and *C. serezkyensis* (Eulipotyphla: Soricidae) // Russian J. Theriol. Vol.23. No.1. P.18–24. doi: 10.15298/rusjtheriol.23.1.02

KEY WORDS: Crocidurinae, *Crocidura*, *Crocidura armenica*, endemic, complete mitochondrial genome, phylogenetic analysis, holotype, zoological collections.

Leonid L. Voyta [leonid.voyta@zin.ru], Laboratory of Theriology, Zoological Institute of the Russian Academy of Sciences, Saint Petersburg 199034, Russia; Tatyana V. Petrova [tatyana.petrova@zin.ru], Valentina A. Panitsina [valentina.panitsina@zin.ru], Semyon Yu. Bodrov [semyon.bodrov@zin.ru], and Natalia I. Abramson [natalia.abramson@zin.ru]: Laboratory of Evolution Genomics and Palaeogenomics, Zoological Institute of the Russian Academy of Sciences, Saint Petersburg 199034, Russia.

Полные последовательности митохондриальной ДНК азиатских эндемичных белозубок: *Crocidura armenica* и *C. serezkyensis* (Eulipotyphla: Soricidae)

Л.Л. Войта*, Т.В. Петрова, В.А. Паницина, С.Ю. Бодров, Н.И. Абрамсон

РЕЗЮМЕ. Белозубки Crocidurinae (Eulipotyphla: Soricidae) — самое крупное подсемейство современных насекомоядных, среди которых род *Crocidura* включает около половины разнообразия семейства. Среди эндемичных *Crocidura* из Центральной и Западной Азии неослабевающий интерес вызывает группа видов "pergrisea". В этом контексте важным вопросом является выяснение таксономического положения и валидность эндемика Армении — *Crocidura armenica*, который с 2014 года исключен из списка видов млекопитающих мира (MSW). В этом исследовании мы собрали полный митохондриальный геном голотипа *C. armenica*. Кроме того, мы получили частичный митохондриальный геном памирского эндемика — *C. serezkyensis*, и белозубки с неясным статусом — *Crocidura* cf. *pergrisea*, из Нахичевани, Азербайджан. Филогенетический анализ митогеномов 28 видов белозубок показывает наиболее близкое положение *C. armenica* и *Crocidura* cf. *pergrisea*, и сестринское к ним положение экземпляра *C. serezkyensis*.

КЛЮЧЕВЫЕ СЛОВА: Crocidurinae, *Crocidura*, *Crocidura armenica*, эндемик, полный митохондриальный геном, филогенетический анализ, голотип, зоологические коллекции.

* Corresponding author

Introduction

The subfamily Crocidurinae unites white-toothed shrews (Eulipotyphla, Soricidae) and is the largest group of living eulipotyphlans (Burgin *et al*., 2018). The genus *Crocidura* Wagler, 1832 within recent shrew fauna covers approximately half of soricids diversity (Burgin & He, 2018). In the context of species diversity of Asian crocidurines, a highly relevant question is the taxonomic position of a local endemic of Armenia *Crocidura armenica* Gureev, 1963, currently represented only by two damaged specimens in the collection of Zoological Institute of the Russian academy of sciences (ZIN): holotype (ZIN 45277) and paratype (ZIN 55321) (Fig. 1). The Armenian shrew was recognized as a taxonomically invalid species (Zaitsev *et al*., 2014) due to limited qualitative and quantitative morphological comparisons of the type specimens, and until now, it has been excluded from the species list of mammal species of the world (Burgin & He, 2018). On the other hand, the ZIN collection includes three undetermined *Crocidura* specimens from Nakhichevan Autonomous Region (Azerbaijan). Zaitsev (1994) identified these specimens as enigmatic rock shrews from the Mountains of West Himalaya (Kashmir): *Crocidura pergrisea* Miller, 1913. Moreover, that author — based on a craniomandibular morphometric dataset and multivariate analysis — supposed that the holotype of *C. armenica* is not conspecific with *C. pergrisea* from Julfa area of Nakhichevan (Zaitsev, 1991). By contrast, we find conspecificity among type specimens of *C. armenica* and specimens from Julfa at least in overall fur and tail coloration, whereas a poor condition of the type specimens' skulls makes correct

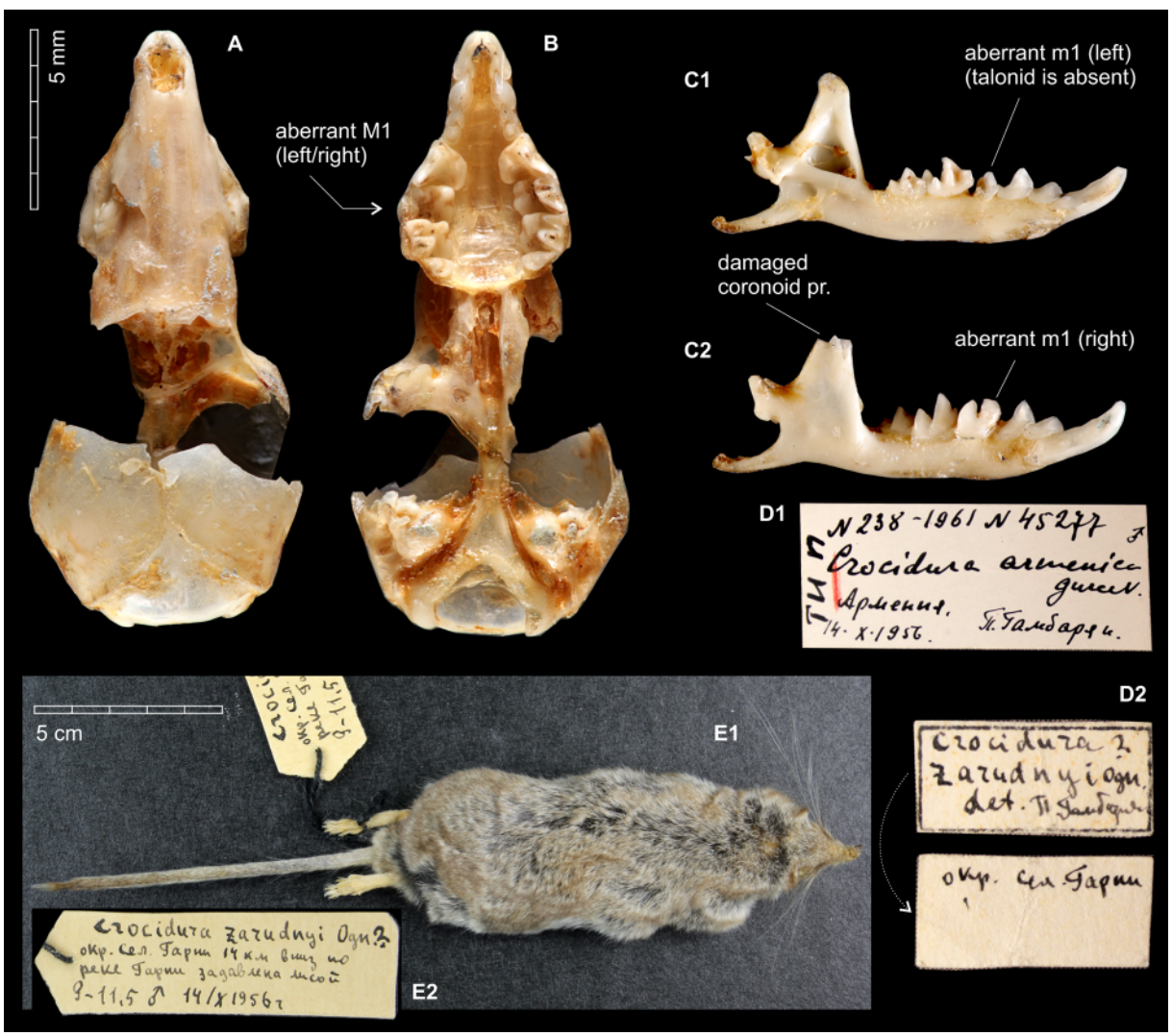

Fig. 1. A skull, hemimandibles, original labels, and overall view of the *Crocidura armenica* holotype (ZIN 45277). A — the heavily damaged skull in dorsal view; B — skull in ventral view; C — left (C1) and right (C2) hemimandibles, in medial and lateral views, respectively; D — labels enclosed with the skull (D1, the main label; D2, a two-sided additional label); (E1) stuffed skin in dorsal view, with a label (E2). The labels are not to scale.

morphological comparisons difficult. In this context, we tried to extract DNA from key museum specimens of *Crocidura* species housed in ZIN collections in an attempt to resolve the important question of taxonomic boundaries of the Armenian shrew.

Materials and methods

The studied specimens are stored at the Laboratory of Theriology of ZIN and labeled as *Crocidura armenica* (holotype), ZIN 45277; *Crocidura serezkyensis*, ZIN 77431; and *Crocidura* cf. *pergrisea*, ZIN 77972 (Table 1).

DNA isolation from museum specimens was carried out in the Laboratory of Evolution Genomics and Palaeogenomics of ZIN in a special room for DNA extraction from museum material (dry skins, etc).

DNA extraction was performed by phenol-chloroform method. Samples were chopped mechanically using scissors, then digested in a lysis buffer with proteinase K at 50°C for 12 hours. DNA was purified by sequential treatment with phenol and chloroform. DNA precipitation was performed using isopropyl alcohol and supersaturated ammonium acetate salt and treated with 70% ethanol. The pellet was dissolved in 40 μL of TE buffer.

Library preparation and sequencing were carried out at the Core Sequencing Centre of Kurchatov Centre for Genome Research (National Research Center Kurchatov Institute, Russia). NGS libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Beijing, China). The resulting PCR products were purified and concentrated using AMPure XP beads (Beckman Coulter, Beverly, MA, USA). The concentration of samples was measured using a Qubit 4 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), while the final quality control of the libraries was implemented using the Bioanalyzer 2100 instrument and the DNA High Sensitivity Kit (Agilent, Boulder, CO, USA). DNA quality was checked with a Qubit 4 fluorometer, and the final distribution of lengths of the libraries' adapter content checking was conducted using Bioanalyzer 2100 (Agilent, Boulder, CO, USA). Paired-end sequencing (2 × 75 bp) was performed on an Illumina HiSeq 4000 system. Bases with a quality score lower than 20 were removed from the sequence reads, and then adaptors were discarded using Trimmomatic (Bolger *et al*., 2014). Clean reads of *C. armenica* were assembled in plasmidSPAdes version 3.10.1 (Bankevich *et al*., 2012; Antipov *et al*., 2016). Clean reads of *C. serezkyensis* and *Crocidura* cf. *pergrisea* (ZIN 77972) were mapped to a reference sequence of *Crocidura shantungensis* Miller, 1901 (GenBank accession No.NC_021398) using Geneious Prime 2019.1 (Biomatters Ltd., Auckland, New Zealand, https://www.geneious.com); fragments with at least 10× coverage were employed to generate a consensus sequence. The contigs were annotated on the MITOS web server (Bernt *et al*., 2013) with default settings. Gene boundaries were checked and refined by alignment against 25 published mitogenome sequences of the *Crocidura* species used for subsequent phylogenetic reconstruction.

Mitochondrial genomes (except the D-loop) of 28 *Crocidura* species (Table 2) — and the Asian house shrew *Suncus murinus* Linnaeus, 1766, serving as an outgroup — were aligned in Geneious Prime 2019.1 via the Geneious alignment with default settings. The phylogenetic reconstruction of the genus *Crocidura* was performed in MrBayes 3.2.6 (Ronquist *et al*., 2012) with the following parameters: nst = mixed and the distribution of the substitution rates between sites. Each analysis involved two replicates with four Markov chains (MCMC) for 1 million generations each, with the results recorded every 1000th generation. Stationarity and convergence of separate runs were assessed using ESS statistics in Tracer v.1.7 (Rambaut *et al*., 2018). A maximum clade credibility tree was constructed based on the trees sampled after 25% burn-in and was visualised in the FigTree v.1.6 software (http://tree.bio.ed.ac.uk/ software/figtree/).

Institutional abbreviations: ISEA, Institute of Ecology and Systematics of Animals (Novosibirsk, Russia);

Table 1. Specimens of *Crocidura* (from the ZIN collection) used in the study. Key: *a catalog number in the Laboratory of Theriology at ZIN; field numbers are given in the brackets; ** a catalog number in the Laboratory of Evolutionary Genomics and Paleogenomics at ZIN; [1]initial species determination by Zaitsev (1991)

Species (age, sex)	Locality	Collector, year	Determined by	Voucher*/ Tissue ID**
Crocidura armenica (mature male)	Garni Village vicinity, Kotayk Province, Armenia; 40.1194° N, 44.7231° E	L. Apoyats, October 14, 1956	Alexey A. Gureev (1963)	ZIN 45277, holotype / ZIN-TER-M-5876
Crocidura serezkyensis (mature female)	Irkht Bay of Sarez Lake, Rushon District of Gorno-Badakhshan Province, Tajikistan; 38.1833° N, 72.6333° E	Mikhail V. Zaitsev, August 07, 1989	Mikhail V. Zaitsev (1991)	ZIN 77431 (56) / ZIN-TER-M-5875
Crocidura armenica [= *Crocidura pergrisea*[1]] (mature female)	Julfa City vicinity, Nakhchivan Autonomous Republic, Azerbaijan; 38.9° N, 45.6167° E	Krjukova and Sevil Radzhabli, November 11, 1972	Boris S. Yudin (confirmed by Zaitsev, 1991)	ZIN 77972 (E-72-27) / ZIN-TER-M-5877

Table 2. The list of *Crocidura* species used for the comparison of mitogenome sequences, with GenBank ID information.

n	Species	GenBank accession No.
1	*Crocidura armenica*, holotype (ZIN 45277)	OR449074
2	*Crocidura armenica* (ZIN 77972)	OR449075
3	*Crocidura serezkyensis* (ZIN 77431)	OR449076
4	*Crocidura attenuata*	NC26204
5	*Crocidura baluensis*	MW815413
6	*Crocidura beatus*	NC27249
7	*Crocidura dongyangjiangensis*	NC56167
8	*Crocidura* ex gr. *foetida-doriae*	MW815412
9	*Crocidura fuliginosa*	NC42762
10	*Crocidura grayi*	NC27247
11	*Crocidura lasiura*	NC29329
12	*Crocidura malayana*	MW815418
13	*Crocidura mindorus*	NC27248
14	*Crocidura negrina*	NC27245
15	*Crocidura nicobarica*	MZ556326
16	*Crocidura ninoyi*	NC27244
17	*Crocidura orientalis*	NC27242
18	*Crocidura palawanensis*	NC27243
19	*Crocidura panayensis*	NC27246
20	*Crocidura russula*	NC56768
22	*Crocidura tanakae*	NC46831
23	*Crocidura wuchihensis*	NC79638
24	*Crocidura* cf. *C. neglecta*	MW815426
25	*Crocidura shantungensis*	NC21398
26	*Crocidura leucodon*	NC70048
27	*Crocidura sibirica*	MH349094
28	*Crocidura* ex gr. *suaveolens-gueldenstaedtii*	ON682400
29	*Crocidura* ex gr. *suaveolens-mimula*	ON682428
30	*Suncus murinus* (outgroup)	NC24604
	Total: 28 species	

NHMW, Natural History Museum Vienna (Austria); USNM, Smithsonian National Museum of Natural History (Washington, USA); ZIN, Zoological Institute of the Russian academy of sciences (St. Petersburg, Russia); ZMMU, Zoological Museum of the Moscow State University (Moscow, Russia).

Results

The complete mitochondrial (mt) genome of *C. armenica* is a closed circular molecule of 16,787 bp in length (Fig. 2) and contains the typical set of 13 protein-coding genes (PCGs), two ribosomal RNA genes (*rrnL* and *rrnS*), 22 transfer RNA (tRNA) genes, and a putative control region. The gene order and organization of the *C. armenica* mt genome are consistent with those of other *Crocidura* mt sequences. Nucleotide composition was found to be considerably biased (A, C, G, and T contents are 32.5%, 24.1%, 13.1%, and 30.3%, respectively) with a GC content of 37.2%. The GC skew of this genome is 0.29. Nine genes (*ND6* and eight tRNA genes) have the reverse orientation, whereas the others are transcribed in the forward direction.

The *C. armenica* mt genome harbors 73 bp overlapping sequences in seven regions. The longest overlap is 43 bp long and located between *ATP8* and *ATP6*. Initiation codons of 13 PCGs of *C. armenica*, except *ND4L* (having GTG), are mostly canonical putative start codons of the ATN type (ATG for *COX1–COX3*, *ND1*, *ND4*, *ND6*, *ATP8*, *ATP6*, and *CYTB*; ATT for *ND3* and *ND5*; and ATC for *ND2*). The typical termination codon (TAA or TAG) occurs in 12 PCGs (in *COX3*, *ND3*, and *ND4*, the TAA stop codon is completed by the addition of 3′ A residues to the mRNA); *CYTB* is terminated with AGA as in other *Crocidura* representatives.

The complete mt genome of *C. armenica* and partial ones of *C. serezkyensis* and *Crocidura* cf. *pergrisea* were submitted to NCBI GenBank under the accession numbers: OR449074, OR449076, and OR449075, respectively.

The result of Bayesian phylogenetic reconstruction by means of the sequences of the 28 species showed that positions of the analyzed specimens of *C. armenica* and *Crocidura* cf. *pergrisea* are the closest within a separate clade, with *C. serezkyensis* as a sister taxon (Fig. 3).

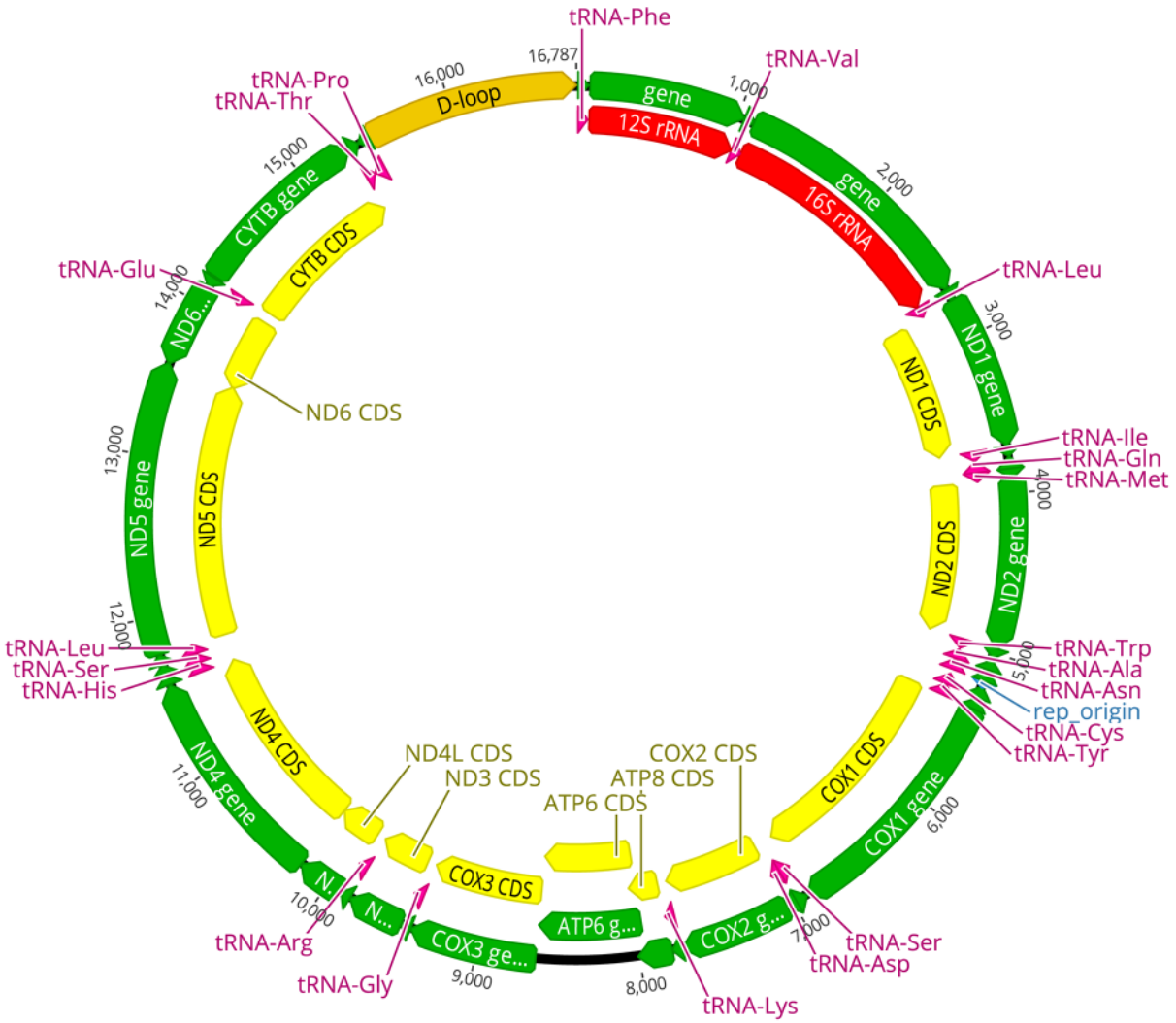

Fig. 2. The map of the mitochondrial genome of *Crocidura armenica* as visualized in Geneious Prime 2019.1.

Discussion and conclusions

This study presents mt sequences of three specimens of rare *Crocidura* species from Central and West Asia. The current results based on the analyses of complete mitagenome sequences support our previous hypothesis about conspecificity of the type specimen of *C. armenica* (holotype) with the analyzed specimen of *Crocidura* cf. *pergrisea* from Julfa. In accordance with this result, we expanded a hypodigm of the Armenian shrew owing to three specimens from Julfa: ZIN 77972 (E-72-27), ZIN 77973 (E-72-28), and ZIN 77976 (E-72-22). In the current paper, these specimens are redefined as *C. armenica*. Conspecificity between analyzed specimen ZIN 77972 and two other specimens was stated in an early karyological study by Grafodatsky *et al*. (1988), which revealed their very unique karyotype features ($2n = 22$, FN = 34) (see Bannikova *et al*., 2023). Unfortunately, in that study, specimen numbers were not marked; nevertheless, we are sure that our specimens match those investigated by those authors. In December 2023, the ZIN collections transferred to the Theriological Collections of the Institute of Ecology and Systematics of Animals (ISEA, Novosibirsk, Russia) a single specimen of *C. armenica* ZIN 77976 (E-72-22) on the basis of a find of its liquid-fixed body. ISEA collections curator Natalia V. Lopatina confirmed the correspondence between Grafodatsky's dataset and our specimens. The Nakhichevan field trip of 1972 was carried out by the ISEA, and after karyological examination, the vouchers were moved to the ISEA collections. After 1988, the specimens have been sent to Mikhail Zaitsev in ZIN collections as "*C. pergrisea*".

In 2023, Bannikova *et al*. (2023) have reported that they analyzed museum DNA of key specimens of the "pergrisea species group" (so-called rocky shrews) from the Collections of the Zoological Museum of the Moscow State University (ZMMU, Moscow, Rus-

sia). Those authors' dataset included nuclear and mt sequences of the three species: *Crocidura arispa* Spitzenberger, 1971 (NHMW 13284, holotype), *Crocidura ramona* Ivanitskaya, Shenbrot et Nevo, 1996 (ZMMU S-165754), and *C. serezkyensis* (ZMMU S-111841, S-111842). Phylogenetic analyses of both mitochondrial and nuclear data revealed a *C. serezkyensis*/*C. arispa* clade, with a sister position of *C. ramona* (Bannikova *et al*., 2023). The sister position of the *C. serezkyensis* specimen (ZIN 77431) toward the clade of *C. armenica* (Fig. 2) allows us to confidently include the Armenian endemic shrew into the "pergrisea species group". It should be noted that name of the group is conventional due to the original data of *C. pergrisea* (holotype USNM 175917) does not revised after the Spitzenberger paper (1971). But, nevertheless, after the Zaitsev (1991) paper the term "pergrisea species group" ordinarily used for the rocky shrews designation.

Our phylogenetic analysis of the complete mt sequences clearly support distinct separation of the "pergrisea species group" from other *Crocidura* species of Asian-tropical, Afrotropic, and Palearctic regions, including species that potentially coexisted such as *C. leucodon* and *C. gueldenstaedtii* (Fig. 2).

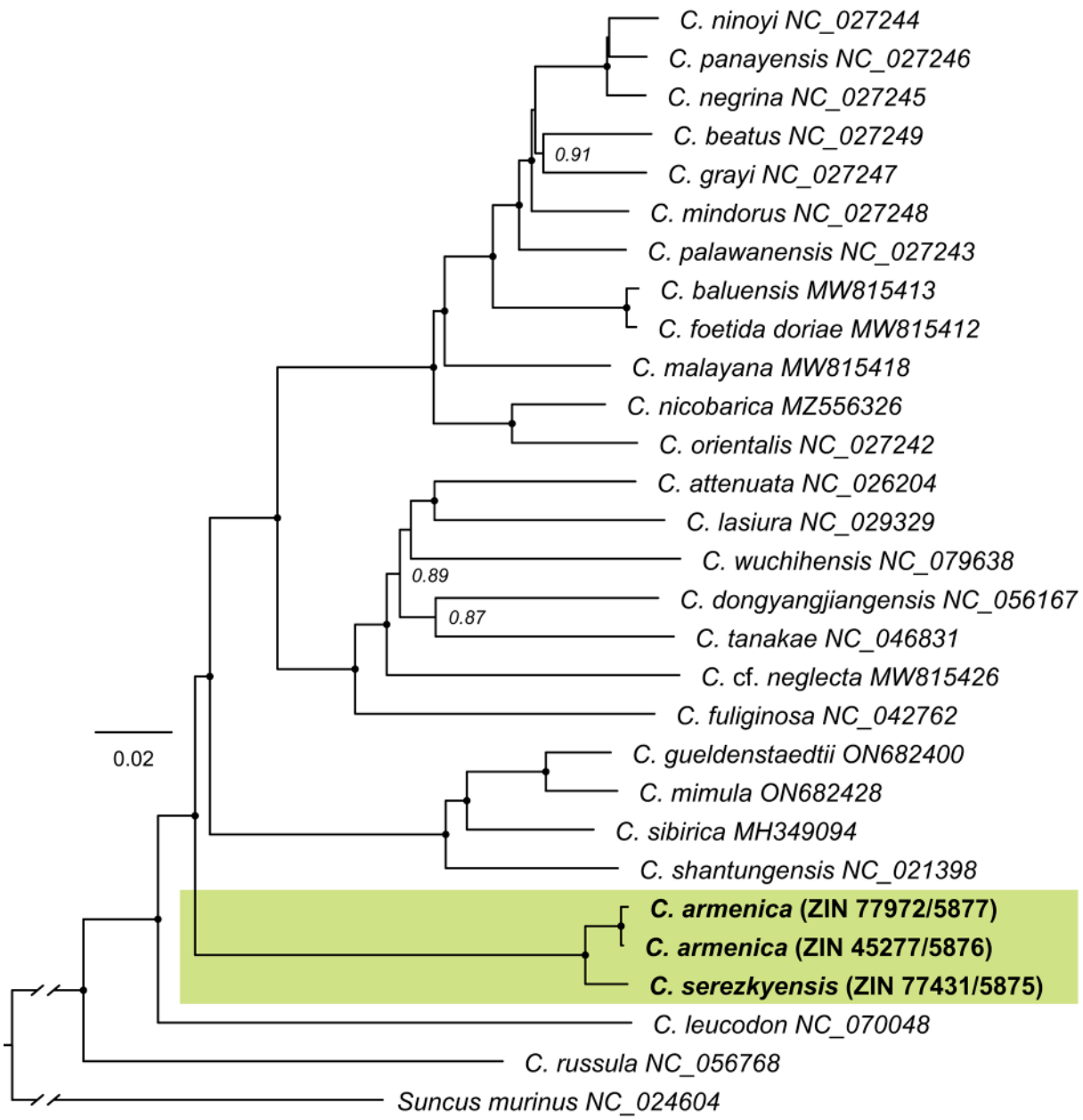

Fig. 3. Bayesian phylogenetic reconstruction of the *Crocidura* genus. Black circles show nodes with 0.98–1.00 Bayesian posterior probabilities (bpp), and node labels indicate bpp values above 0.7 and below 0.98. Sequences obtained in the current study are highlighted in bold.

Currently, two questions remain open. First is a conspecificity between the holotype (analyzed ZIN 45277) and paratype (ZIN 55321) of *C. armenica*. The skull of paratype is also heavily damaged. Mikhail Zaitsev has excluded it from the morphometric analyses because of the skull's condition (Zaitsev, 1991). Our attempt at extraction of the museum DNA from the paratype was unsuccessful due to presumed formaldehyde in the preservation liquid of the specimen. Second is a particular position of *C. armenica* in a combined phylogenetic analysis including other species of the "pergrisea species group" is also unclear. Until such analysis is carried out, we are not able finally resolve the issue of the taxonomic validity of *C. armenica*.

ACKNOWLEDGMENTS. The study partly used the collection materials of the Zoological Institute of the Russian Academy of Sciences (Saint Petersburg, http://www.ckp-rf.ru/usu/73561/). The study was completed within the framework of the Federal Themes of the Zoological Institute No. 122031100282-2. Funding for technical part of the study was provided by the Ministry of Science and Higher Education of the Russian Federation, project 075-15-2021-1069. The authors are grateful to collection keepers of the Laboratory of Theriology at ZIN, Olga Makarova and Eugenia Maksimova, for assistance with the tissue extraction. We are very indebted to our colleague Dr. Olga V. Bondareva for a discussion of the results and many valuable comments. Special thanks to Nikolai Shevchuk (http://shevchuk-editing.com/) for language editing. The authors are gratitude two anonymous reviewers for the substantial comments and the text correction.

References

Antipov D., Hartwick N., Shen M., Raiko M., Lapidus A. & Pevzner P.A. 2016. PlasmidSPAdes: Assembling Plasmids from Whole Genome Sequencing Data // Bioinformatics. Vol.32. P.3380–3387.

Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., Nikolenko S.I., Pham S. & Prjibelski A.D. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing // Journal of Computational Biology. Vol.19. P.455–477.

Bannikova A.A., Lisenkova A.A., Solovyeva E.N., Abramov A.V., Sheftel B.I., Kryštufek B. & Lebedev V.S. 2023. The first phylogenetic data on the elusive shrews of the *Crocidura pergrisea* species complex // Hystrix. Vol.34. P.33–38.

Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf M. & Stadler P.F. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation // Molecular Phylogenetics and Evolution. Vol.69. P.313–319.

Bolger A.M., Lohse M. & Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data // Bioinformatics. Vol.30. P.2114–2120.

Burgin C.J., Colella J.P., Kahn P.L. & Upham N.S. 2018. How many species of mammals are there? // Journal of Mammalogy. Vol.99. P.1–14.

Burgin C.J. & He K. 2018. Family Soricidae // Wilson D.E. & Russell A.M. (eds.). Handbook of the Mammals of the World. Vol.8. Insectivores, Sloths and Colugos. Barcelona: Lynx Edicions. P.332–551.

Grafodatsky A.S., Radzhabli S.I., Sharshov A.V. & Zaitsev M.V. 1988. Karyotypes of five *Crocidura* species of the USSR fauna // Tsitologiya. Vol.30. P.1247–1250 [in Russian, with English summary].

Rambaut A., Drummond A.J., Xie D., Baele G. & Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7 // Systematic Biology. Vol.67. No.5. P.901–904.

Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A. & Huelsenbeck J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space // Systematic Biology. Vol.61. P.539–542.

Spitzenberger F. 1971. Eine neue, tiergeographish bemerkenswerte *Crocidura* (Insectivora, Mammalia) aus der Türkei // Annalen des Naturhistorischen Museums in Wien. Serie B für Botanik und Zoologie. Vol.75. P.539−552.

Zaitsev M.V. 1991. Species composition and questions of systematics of white-toothed shrews (Mammalia, Insectivora) of the fauna of USSR // Proceedings of the Zoological Institute, USSR Academy of Sciences. Vol.243. P.3–46 [in Russian, with English summary].

Zaitsev M.V., Voyta L.L. & Sheftel B.I. 2014. [The Mammals of Russia and Adjacent Territories. Lipotyphlans]. Saint Petersburg: Nauka. 391 p. [in Russian].