ORIGINAL ARTICLE

Numerous insertions of mitochondrial DNA in the genome of the northern mole vole, *Ellobius talpinus*

Kristina Kuprina[1,2] · Antonina Smorkatcheva[2] · Anna Rudyk[2] · Svetlana Galkina[3]

Abstract

Background *Ellobius talpinus* is a subterranean rodent representing an attractive model in population ecology studies due to its highly special lifestyle and sociality. In such studies, mitochondrial DNA (mtDNA) is widely used. However, if nuclear copies of mtDNA, aka NUMTs, are present, they may co-amplify with the target mtDNA fragment, generating misleading results. The aim of this study was to determine whether NUMTs are present in *E. talpinus*.

Methods and results PCR amplification of the putative mtDNA *CytB*-D-loop fragment using 'universal' primers from 56 *E. talpinus* samples produced multiple double peaks in 90% of the sequencing chromatograms. To reveal NUMTs, molecular cloning and sequencing of PCR products of three specimens was conducted, followed by phylogenetic analysis. The pseudogene nature of three out of the seven detected haplotypes was confirmed by their basal positions in relation to other *Ellobius* haplotypes in the phylogenetic tree. Additionally, 'haplotype B' was basal in relation to other *E. talpinus* haplotypes and found present in very distant sampling sites. BLASTN search revealed 195 NUMTs in the *E. talpinus* nuclear genome, including fragments of all four PCR amplified pseudogenes. Although the majority of the NUMTs studied were short, the entire mtDNA had copies in the nuclear genome. The most numerous NUMTs were found for *rrnL*, *COXI*, and D-loop.

Conclusions Numerous NUMTs are present in *E. talpinus* and can be difficult to discriminate against mtDNA sequences. Thus, in future population or phylogenetic studies in *E. talpinus*, the possibility of cryptic NUMTs amplification should always be taken into account.

Keywords NUMTs · Mitochondrial pseudogenes · Heteroplasmy · Arvicolinae · Rodentia

Introduction

The presence of sequences with significant homology to mitochondrial DNA (mtDNA) has been established in the nuclear genome of many eukaryotes [1, 13, 16, 25, 33, 34, 50]. These sequences (*Nu*clear *mit*ochondrial sequences or NUMTs, also called mtDNA pseudogenes) can be of various

amounts and lengths, and most of them are non-functional degenerate copies of various mtDNA fragments [6]. Being numerous and predominantly selectively neutral, these "relics of ancient DNA" are valuable molecular markers in evolutionary biology that can be used in rooting and resolving phylogenetic reconstructions and estimating nuclear mutation rates [5].

Information on the distribution of NUMTs integration sites has been accumulating (e.g., [18, 21]). In organisms with well-developed genome assemblies, NUMTs can be identified bioinformatically through targeted searching for mtDNA inserts in nuclear genome sequences. At the early stage of genome sequencing and assembly studies, NUMTs are detected occasionally as artefacts, when nuclear DNA is co-amplified in PCR using mtDNA-specific primers [7, 8, 11, 22, 32]. Therefore, from a practical point of view, NUMTs are a potential source of contamination during PCR amplification of mtDNA fragments. The risk of pseudogene co-amplification further increases with 'universal' primers.

✉ Kristina Kuprina
   kristina.kuprina@uni-greifswald.de

1 Institute of Botany and Landscape Ecology, University of Greifswald, Soldmannstr. 15, Greifswald 17489, Germany

2 Department of Vertebrate Zoology, Saint Petersburg State University, Universitetskaya nab. 7/9, Saint Petersburg 199034, Russia

3 Department of Genetics and Biotechnology, Saint Petersburg State University, Universitetskaya nab. 7/9, Saint Petersburg 199034, Russia

NUMTs are known to be a source of errors in phylogeny and phylogeography reconstruction [2, 20, 27, 30, 39, 41, 45] DNA barcoding [8, 44], and disease association with mutations in mtDNA [46, 48]. Analyse of mtDNA sequence diversity is a popular tool for assessing population structure and dynamics, while differentiation of an authentic mitochondrial sequence from its pseudogene is critical for the correct interpretation of haplotype diversity. Therefore, identification of NUMTs in the genomes of the species used as models in population studies is of great importance.

Mole voles (*Ellobius*, Arvicolinae, Cricetidae) are highly specialised subterranean rodents inhabiting the grasslands of Eurasia. These rodents represent an attractive model in population ecology studies due to their lifestyle, sociality and unusual life-history traits, e.g., slow growth and development, delayed sexual maturity, and longevity extreme among voles [19, 22, 31, 37]. However, until nowadays, only *cytochrome B* (*CytB*) sequences were investigated in several population genetic studies of this genus [12, 15, 43]. Meanwhile, the D-loop region has several advantages as a molecular marker for assessment of genetic diversity and structure: it is non-coding (therefore, we can presume they are selectively neutral) and the most variable part of mtDNA, mainly due to mutations rather than recombination [17, 40, 42]. During the development of a primer combination to amplify a fragment of D-loop region for the northern mole vole (*Ellobius talpinus*), we detected multiple occurrences of double peaks in the chromatograms. We inspected cloned PCR products via sequencing and phylogenetic analysis to test the hypothesis that the nuclear genome of the northern mole vole does contain NUMTs homologous to D-loop region and *CytB* fragments. Additionally, the genome assembly of *E. talpinus* was analysed to identify other NUMTs for use in evolutionary studies and checking their potential interference against the existing studies of mitochondrial genetic variation.

Materials and methods

The northern mole voles (*Ellobius talpinus*) were live-trapped in the Novosibirsk (n = 56) and Chelyabinsk regions (n = 1) of Russia as part of population studies (Fig. 1). Three Zaisan mole voles (*E. tancrei*) from southwestern Tajikistan were also included in this study. The animals were marked by toe clipping, and the phalanges were further stored in 96% ethanol for genetic analysis. All applicable institutional guidelines for the care and use of animals were followed. After the toe clipping, animals were released into a burrow where they were captured. The study was conducted under the ethical clearance from the Ethical Committee of Saint Petersburg State University (Statement no. 131-03-9 issued on 22 November, 2021) in accordance with the National Research Council (2011).

DNA was extracted according to the standard phenol-chloroform procedure [36]. For PCR, a pair of 'universal' primers ArvF 5'-GCCTCAATCGCCTACTTCAC-3' and ArvR 5'-GGCTGATTAGACCCGTACCA-3' was designed by aligning the published mtDNA sequences of Arvicolinae (NCBI: FJ502319, JX996084, JX996089, AF192739, AF348371 - AF348384, AF267284, HQ395088, DQ198847, GQ452102). The desired fragment was expected to be 670 bp
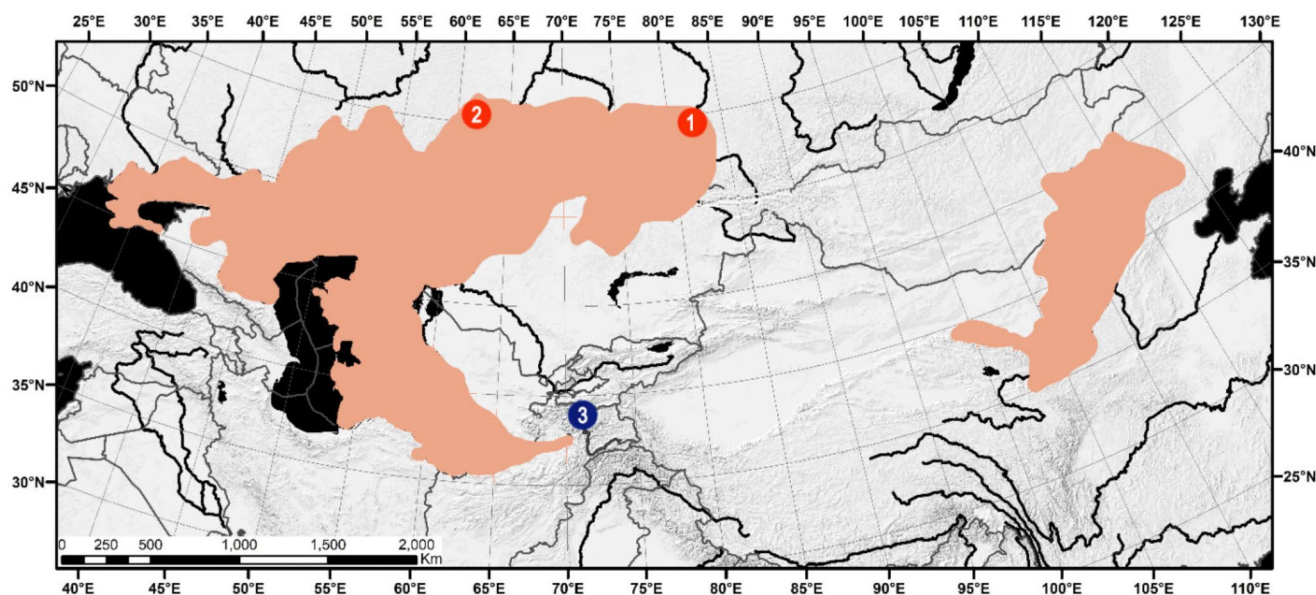

Fig. 1 Species distribution map of *Ellobius talpinus* with sampling sites: 1 - Novosibirsk population (n = 56); 2 - Chelyabinsk population (n = 1), and 3 - sampling site of *E. tancrei* (n = 3). Map (modified) from [23]

long and include 84 bp of *CytB* gene, two tRNA genes, and a part of D-loop which comprised the first hypervariable region (Supplementary Fig.1). PCR with these primers was conducted for all *Ellobius* specimens. PCR reaction mixture (20 μl) contained 1X Taq-buffer, 1 U HotStart Taq polymerase, 1.5 mM MgCl$_2$ (all from Axon Labortechnik, Germany), 4 μM dNTPs, 0.2 μM each primer, and 60 ng of DNA as a template. Amplification was done in a thermocycler MJ Mini (Bio-Rad Laboratories, USA) according to the following program: (i) 13 min at 95 °C; (ii) 35 cycles of amplification consisting of 45 s at 95 °C, 30 s at 55 °C and 30 s at 72 °C; (iii) 3 min final extension at 72 °C. All PCR products including no template controls were analysed by agarose gel electrophoresis. For three *E. talpinus* from the Novosibirsk region, PCR products were purified on silica magnetic beads (Sileks, Russia) and cloned using the "InsTAclone PCR Cloning Kit" (Thermo Fisher Scientific, USA). PCR amplicons and cloned insert fragments were sequenced on an ABI Prism 3500xl analyser (Thermo Fisher Scientific, USA). All laboratory procedures were conducted using filter tips; no-template controls were included in every PCR and sequenced. To additionally verify the absence of contamination, 8 samples of *E. talpinus* were extracted and amplified with new reagents in a different laboratory.

Sanger sequencing chromatograms were visually inspected using Chromas 2.6.6 (Technelysium Pty Ltd, Australia), and in the case of double peaks, the highest peak was used to determine the nucleotide. Forward and reverse reads were compared pairwise to exclude the possible sequencing errors. All obtained sequences were manually aligned using MEGA X [24], and the resulting haplotypes were checked for specificity using BLASTN. MEGA X software was also used for checking the reading frame in a fragment of *CytB*.

To check whether NUMTs had been co-amplified, a phylogenetic analysis of all obtained haplotypes was conducted. When phylogenetic relationships within a taxon are well known, pseudogenes can often be detected by the atypical length of their branches and irregular topology [3, 4, 9, 38], since the mutation rate in NUMTs is about ten times lower compared to mtDNA [5, 26]. D-loop sequences of *E. tancrei*, steppe lemming, *Lagurus lagurus*, yellow steppe lemmings, *Eolagurus luteus*, and European water vole, *Arvicola amphibius*, were included in the analysis; the large-eared vole, *Alticola macrotis* and the bank vole, *Myodes glareolus* were used to root the tree. The sequences for *E. tancrei* were obtained in this work using 'universal' primers, the rest of the sequences were retrieved from GenBank (Supplementary Table 1). We used 47 bp fragment of *cytochrome B*, *tRNA-Thr*, *tRNA-Pro*, and 450 bp D-loop fragment, in total 631 bp corresponding to positions 15,212–15,842 bp of the *E. talpinus* mtDNA (NCBI: NC_054160). The best-fit substitution model (HKY+G) was found using MEGA X.

Accordingly, the combination of nst = 6, a discrete Gamma distribution (+G) with 4 rate categories with assuming that a certain fraction of sites is evolutionarily invariable were used for Bayesian analysis. The phylogenetic tree was built using MrBayes v.3.2 [35] (10,000,000 generations, 25% burn-in) and visualised in FigTree v.1.4.3 (tree.bio.ed.ac.uk/software/figtree/).

The presence of pseudogenes in the nuclear genome of *E. talpinus* was verified by mapping sequencing reads from the SRR3497471 NCBI SRA database to the sequence of each of six haplotypes. At first we assembled a mitogenome for sample SAMN04317029 (www.ncbi.nlm.nih.gov/biosample/SAMN04317029), and we excluded all mitochondrial reads from the subsequent analysis. Mapping the remaining reads to haplotypes was performed using the 'Map to reference' option in Geneious Prime (www.geneious.com/) with strong selectivity (maximum 1% of gaps and 1% of mismatches per read were allowed). We selected split reads where a part of the read is mapped to the haplotype and the other part half to the nuclear genome. Such split reads were used in BLAST search to identify the specific contigs in the NCBI whole-genome sequencing (WGS) database GCA_001685095.1 corresponding to the reference genome ETalpinus_0.1 (www.ncbi.nlm.nih.gov/datasets/taxonomy/329620/) and to define the breakpoints. The presence of other NUMTs in the *E. talpinus* nuclear genome was evaluated through BLASTN search on the only available genome assembly GCA_001685095.1. We split *E. talpinus* mtDNA reference sequence (NCBI: NC_054160) into 16 non-overlapping segments of 1000 bp plus one fragment of 637 bp, and aligned them against GCA_001685095.1. All contigs found were additionally checked for the presence of NUMTs through alignment to NC_054160 using the 'Align whole genomes' option (Mauve chromosome alignment algorithm) implemented in Geneious Prime software. Annotation of contigs, including NUMTs lengths, their start and end positions, and the corresponding sites in the *E. talpinus* mtDNA NC_054160 was also done in Geneious Prime (Supplementary Table 2).

Results

Using 'universal' primers ArvF and ArvR, we amplified the PCR products from 56 Novosibirsk samples and one Chelyabinsk sample of *E. talpinus*. All amplicons produced single bands in electrophoresis. They were sequenced, and four haplotypes (A, B, C, and D) were detected. The haplotypes A, C and D had 1 (A vs. C), 2 (A vs. D) or 3 (C vs. D) variable sites; however, haplotype B differed from A, C and D in 33–35 positions. Most of the Novosibirsk samples (64%) as well as the sample from the very distant Chelyabinsk region

had haplotype B. For a number of samples, the sequencing resulted in chromatograms containing double peaks at more than 30 nucleotide sites. Clean laboratory practice combined with absence of PCR products in the agarose gel (Supplementary Fig. 2a) and in the chromatograms for no-template controls confirms the absence of contamination; however, double peaks of different heights (Supplementary Fig. 2b) in > 90% of the working chromatograms persisted, indicating potential NUMT co-amplification.

Following the demonstration of abnormally diverged sequences generated by 'universal' primers, we cloned the PCR products of three Novosibirsk samples and sequenced several (3, 5, and 16) plasmids per individual. No double peaks were found in the chromatograms. We identified haplotypes A, B, C and D, and three additional ones: E, F, and G. Haplotype B was found in all three samples along with either A, or C, or D. Haplotypes E, F and G differed from haplotype A in 89, 62 and 66 bp, respectively. The BLAST search of the haplotypes found the closest similarity with the mtDNA of *E. talpinus*. In the *CytB* fragment, neither an additional stop codon nor any frame shifts were detected for any of the haplotypes. All haplotypes obtained by 'universal' primers were uploaded to GenBank (NCBI: OR662053 (*E. tancrei*), OR662050 - OR662052 (*E. talpinus* haplotypes

A, C, and D), OR662054 - OR662057 (*E. talpinus* pseudogenes B, E, F, and G)).

Phylogenetic analysis of the detected *Ellobius* haplotypes together with corresponding sequences of the selected arvicolines revealed the basal position of *E. talpinus* haplotypes E, F, and G relative to the other haplotypes of the genus Ellobius (Fig. 2). The remaining *E. talpinus* haplotypes together with *E. tancrei* formed a monophyletic group.

To check whether a primer with greater specificity can decrease or eliminate co-amplification of revealed pseudogene(s), a new primer, EtalpF (5'-TCAAGAAGGAAGGACCTACCC-3'), was designed based on the results of phylogenetic analysis and mitogenome of *E. talpinus* (NCBI: NC_054160.1). This primer, together with ArvR, was used to amplify mtDNA D-loop of 16 Novosibirsk samples, which previously showed haplotype B. All of the obtained sequences were unambiguous. Haplotype B was found in two samples, while the rest had haplotypes A or D (haplotype C was not detected due to a shorter size of the amplified fragment).

To test bioinformatically whether the sequences corresponding to putative pseudogenes are present in the nuclear genome, we mapped the sequencing reads from the *E. talpinus* SRR3497471 NCBI SRA database to each of the six revealed haplotypes. After exclusion of all mitochondrial
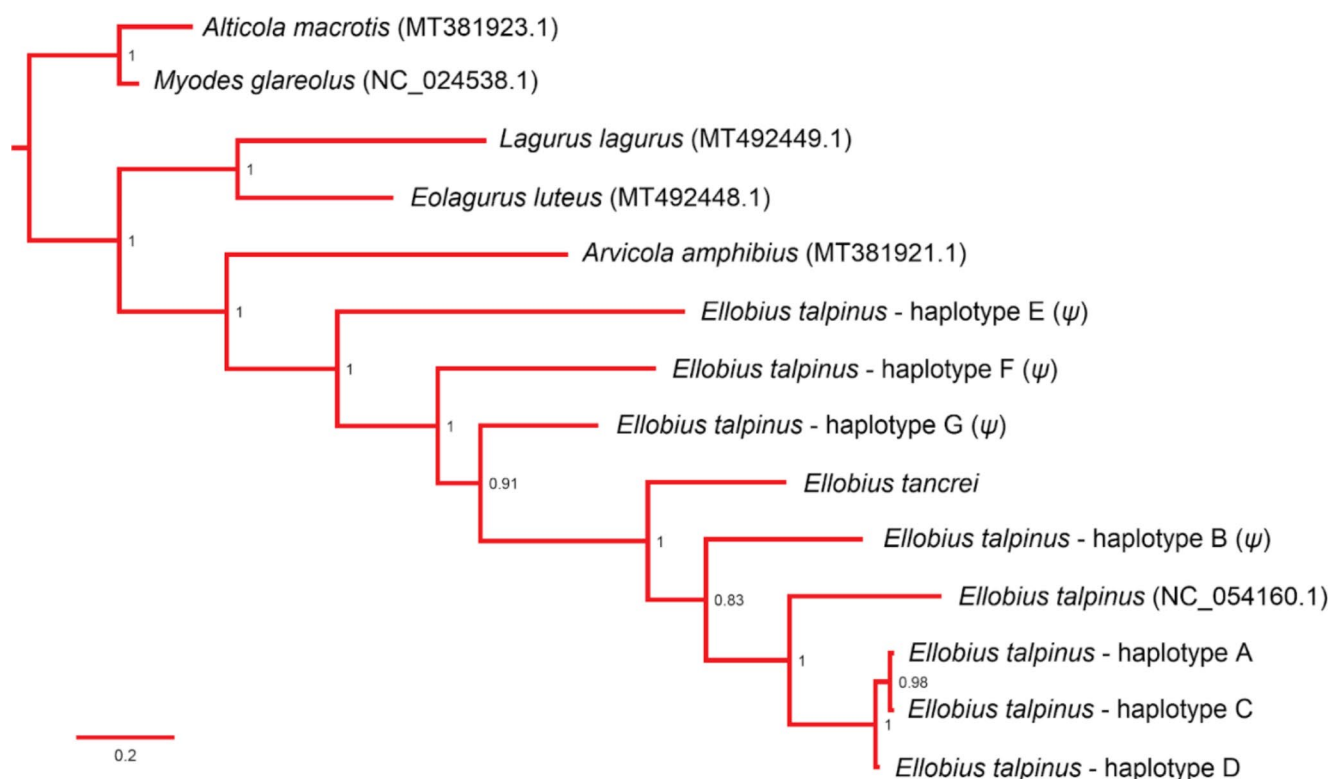

Fig. 2 Bayesian tree of 7 Arvicolinae species based on the 631 bp fragment of mtDNA, including the haplotypes of *E. talpinus* and *E. tancrei*, amplified with 'universal' primers. Putative pseudogenes are indicated by ψ. The sequences not obtained in this study are shown with the NCBI accession numbers in brackets

reads, no additional reads aligned to haplotypes A, C, and D. As for haplotypes B, E, F, and G, we identified split reads, which allowed the detection of several contigs in the *E. talpinus* GCA_001685095.1 genome database. When searching for 1 kb fragments of mtDNA sequences in the GCA_001685095.1, we found 240 WGS contigs (Supplementary Table 2); 45 were found to be entirely mitochondrial; they were annotated but excluded from the subsequent counting (Supplementary Table 2). In the remaining 195 contigs, the boundaries of a mtDNA fragment copied in a NUMT were determined by the alignment between the corresponding contig and the NC_054160.1 mitogenome (Fig. 3a, Supplementary Table 2). NUMTs ranged in size from 31 to 4536 bp, median being 315 bp (Supplementary Fig.4). The majority of NUMTs were short insertions: 65.6% were less than 500 bp in size. NUMTs for *ND5*, *CYTB*, *COXI* protein-coding genes as well as for D-loop and both rRNAs were the most copious. In contrast, NUMTs for *ND4L* and *ND3* genes were rare (Fig. 3b).

Discussion

In this study, we present two arguments supporting the hypothesis that the nuclear genome of *E. talpinus* contains numerous insertions of various mtDNA fragments. The first argument comes from a practical problem related to amplification of D-loop sequences using 'universal' primers for the northern mole vole. The frequent detection of double peaks in the *CytB*-D-loop fragment chromatograms and the presence of 'haplotype B' in animals from very distant sampling sites (located more than 1300 km from each other, Fig. 1) gave rise to the NUMTs co-amplification hypothesis. Molecular cloning of PCR-amplicons supported this hypothesis indicating the presence of haplotype B, together with one of three other previously revealed haplotypes (A, C, or D) in each individual. Moreover, cloning discovered three extra haplotypes (E, F, and G) whose pseudogene nature was confirmed by their basal positions in relation to the other obtained *Ellobius* haplotypes in the phylogenetic tree (Fig. 2) and by finding sequencing reads comprising both mitochondrial and non-mitochondrial DNA fragments (Supplementary Fig.5). The second evidence comes from the results of our BLAST search in *E. talpinus* GCA_001685095.1 genome database showing that 195 WGS contigs contain sequences of mitochondrial origin of various lengths. Thus, the genome of the northern mole vole is not an exception among other voles in terms of the NUMT content [1, 3]. This is the first report on NUMTs in the nuclear genome of *Ellobius*.

Some cases of predominant amplification of pseudogenes by universal primers were described in the past [11, 49]. In our case, the conservative primer ArvF designed using the sequences of other Arvicolinae species had two non-specific to *E. talpinus* sites. These sequences lead to amplification of haplotype B with high probability, and, in a minority, haplotypes E, F, and G. In this work, in order to minimise the presence of pseudogenes in the PCR products, we designed a new primer, EtalpF, which considerably diminished, although did not completely eliminate, the amplification of the pseudogene. In a recently initiated population genetic study of mole voles from the Saratov Region, we obtained 75 sequences using the EtalpF primer (Rudyk et al., in preparation). Several chromatograms showed double peaks at certain nucleotide sites; their positions indicated co-amplification of pseudogene B together with the authentic D-loop region fragment. Moreover, one of the unambiguous sequences completely corresponded to pseudogene B. To discriminate between pseudogenes and orthologous coding mitochondrial sequences, the reading frame and stop codons can be inspected. However, this method is unsuitable for detecting pseudogenes of the D-loop region because of the non-coding nature of the fragment. So, we strongly urge that when analysing mitochondrial polymorphism in these rodents, the possibility of cryptic NUMTs amplification should always be kept in mind.

When double peaks on chromatograms and/or unusually common haplotype(s) are detected for D-loop region sequences, checking the phylogenetic relationships of all identified haplotypes in order to detect and exclude haplotypes with an anomalous basal position has been recommended [2, 4, 7, 9]. Our phylogenetic reconstruction of seven haplotypes of *E. talpinus* (obtained with the "universal" primers) along with the corresponding fragments for six other arvicoline species showed that three haplotypes of *E. talpinus* (E, F, and G) had a basal position relative to the divergence of two *Ellobius* species, indicating the early origin and nuclear source of sequenced fragments (Fig. 2). The position of haplotype B in the *Ellobius* clade did not receive high support, but was basal relative to the clade encompassing haplotypes A, C, D, and *E. talpinus* NC_054160.1. This basal position of haplotype B together with its wide geographical distribution and most frequent occurrence indicate the ubiquitous and pseudogenic nature of haplotype B. The aligning of haplotypes B, E, F, and G with the reads from the *E. talpinus* nuclear genome supported our characterization of them as NUMTs.

In our work, we identified 195 NUMTs by aligning the *E. talpinus* mitochondrial genome on the nuclear genome using BLASTN (Supplementary Table 2). The majority of NUMTs were short insertions consistent with ongoing selection against large NUMTs [10]. According to our estimation, the absolute and relative to the nuclear genome amount of NUMTs in the northern mole vole is 125.03 kb or 0.004%.
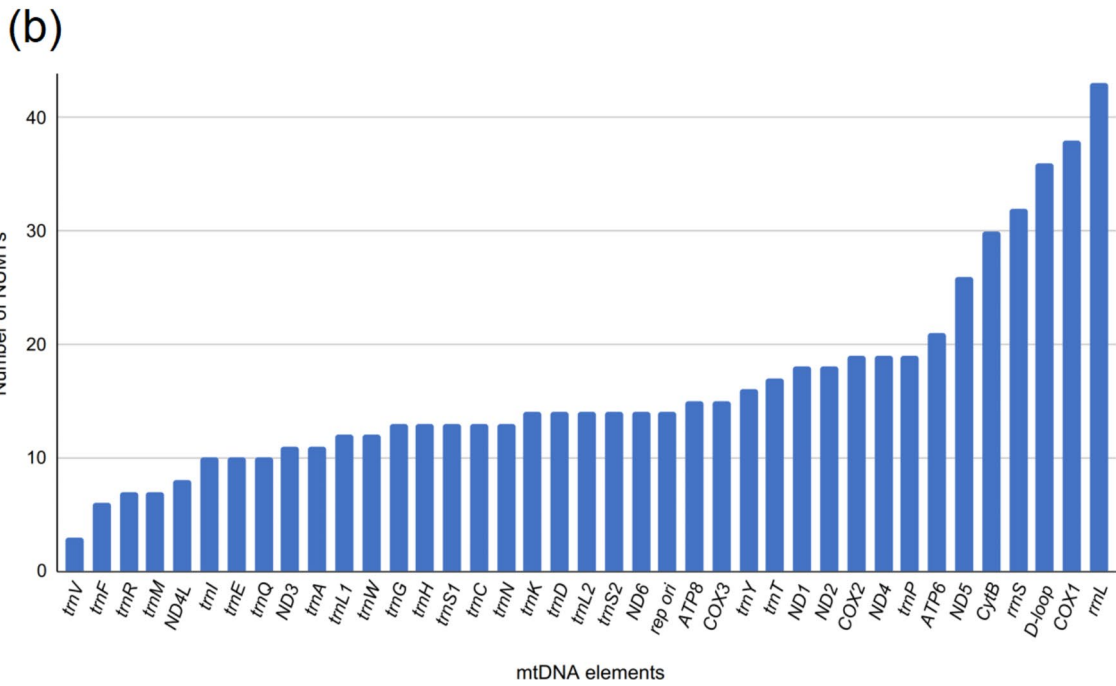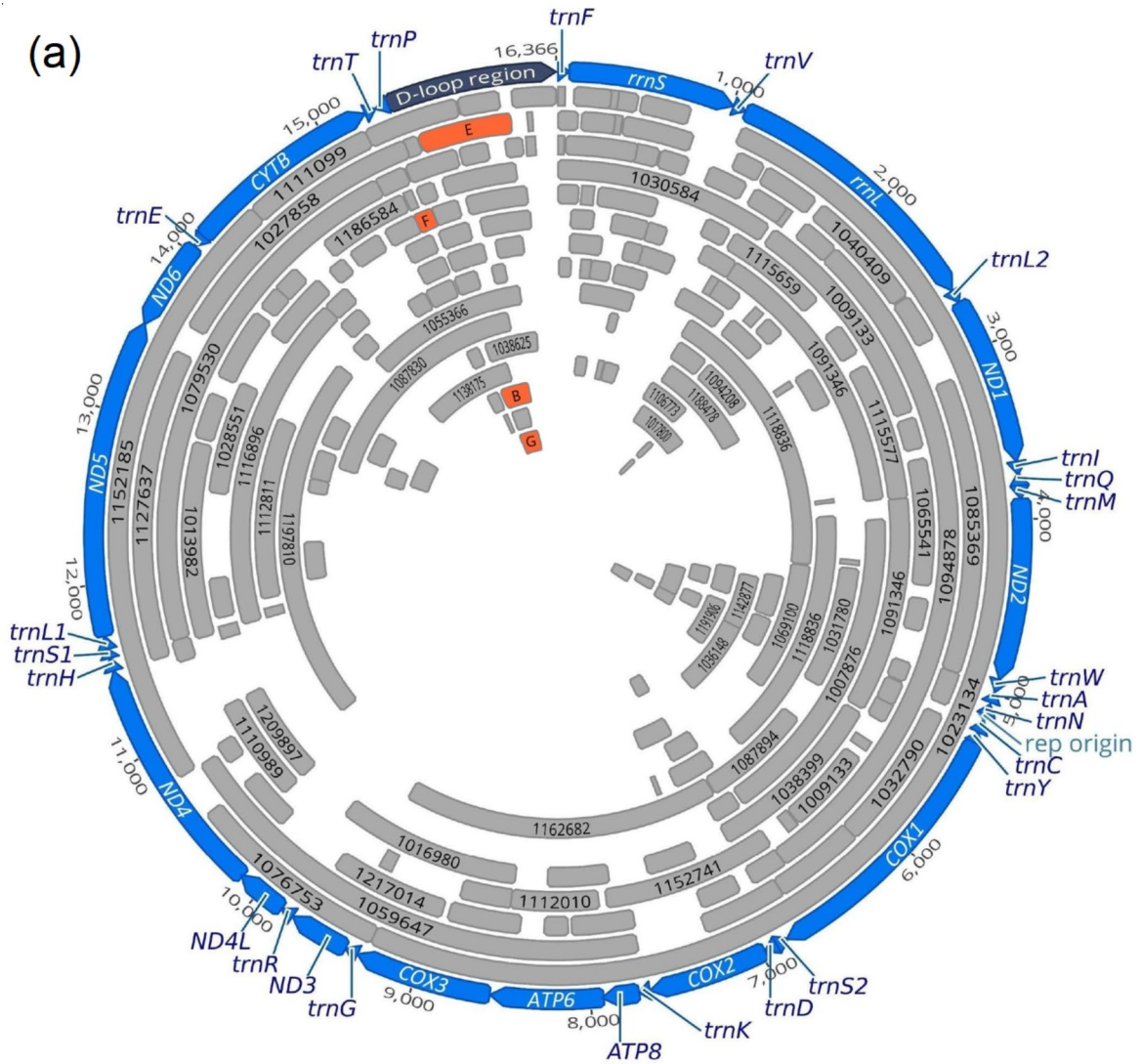
Fig. 3 Distribution of NUMTs corresponding to various elements of the *E. talpinus* mitochondrial genome. (a) Circular representation of *E. talpinus* mtDNA (blue circle), which contains the genes for 13 energy pathway proteins, two rRNAs, and 22 tRNAs, and the replication origin and D-loop control region. In mtDNA circle, the grey lines indicate NUMT insertions found in WGS contigs from the GCA_001685095.1 genome database. Letters E, F, B and G label the contigs containing fragments of pseudogenes E, F, B, G, correspondingly. The length of each line reflects the length of the corresponding NUMT. The longest NUMTs are marked with the numbers of the corresponding contigs. (b) The number of contigs bearing NUMTs for each of the 39 mtDNA elements including replication origin element and D-loop region. Each contig was counted even if it contained a fragment of an element

This is two times more than found in another rodent, *Mus musculus* (Murinae, Muridae) (37.67 kb, 0.002%) [6]. This discrepancy can be explained by the insufficient completeness of the available *E. talpinus* genome used in the study (number of contigs 350,460, N50 = 15.2 kb, www.ncbi.nlm.nih.gov/datasets/genome/GCA_001685095.1/). At the same time, among the rodent genomes, the median number of NUMTs ranges from 168 to 11,930 (median 477), being in the Arvicolinae voles 625 (*Microtus agrestis*), 334 (*M. ochrogaster*), 2,952 (*M. arvalis*), 290 (*Arvicola amphibius*), and 502 (*Myodes glareolus*) [1]. Thus, our modest 195 NUMTs found in *E. talpinus* may be considered a reliable number for these small rodents. Further improvement of the northern mole vole genome assembly will undoubtedly clarify the situation.

We demonstrated that in *E. talpinus*, the entire mtDNA is involved in NUMTs. The most numerous NUMTs were found for *ND5*, *CYTB*, *COXI*, *rrnL*, *rrnS*, and D-loop region; *ND4L*, *ND3* and genes for some tRNAs seldom produce NUMTs (Fig. 3a). Interestingly, *CYTB*, D-loop region, *rrnS* and *rrnL* elements are adjacent, while ND4L and ND3 are situated at the opposite pole (Fig. 3a). None of the previous studies appear to have found local transfer preferences across the mitochondrial genome in rodents [1, 2]. However, in humans, mtDNA breakpoints were found to be more common to the non-coding D-loop region compared to other sites, and were less likely to involve ND3 and ND4L genes [10].

The issue of the D-loop region contribution to NUMTs is still under discussion. Mourier et al. [29] observed a deficiency in the D-loop region derived NUMTs (in humans) but attributed that to the difficulty of NUMT detection in this rapidly evolving region. At the same time, mtDNA D-loop fragments were found in the genomes of rodent species [3]. The presence of D-loop derived sequences in the nuclear genome is indicative of DNA-mediated rather than RNA-mediated NUMT insertions, since the control region has no intermediate RNAs [14, 47]. The current plausible explanation of the emergence of mitochondrial pseudogenes involves mtDNA transcription and associated replication occurring in the D-loop region. This is in line with the recent description of mitochondrial apoptosis mechanism mediated by the BCL-2 family proteins BAK and BAX inducing outer mitochondrial membrane permeabilization through large pores formation [28]. These BAK/BAX macropores allowed the inner mitochondrial membrane to penetrate into the cytosol, taking with it the components of the mitochondrial matrix, including the mitochondrial genome. Once in the cytoplasm, mitochondrial DNA can, under certain circumstances, integrate into the nuclear genome. Although it is unlikely that NUMTs are functional per se due to differences between the nuclear and mitochondrial genetic codes [5], it can be assumed that in rare cases they may perform a functional role as regulatory elements or through the creation of new exons.

Conclusion

Multiple NUMTs that are difficult to discriminate against mtDNA sequences were detected in the genome of *E. talpinus*. A combination of primers for preferential amplification of mitochondrial fragment of the *CytB*-D loop region has been developed for the future population genetic studies of this species. For any mitochondrial genetic marker, it is strongly recommended to mind the possibility of cryptic NUMTs amplification in *E. talpinus*.

Declarations

Ethics approval The procedures related to manipulation of animals were approved by the Ethical Committee of Saint Petersburg State

University (Statement no. 131-03-9 issued on 22 November, 2021) in accordance with the National Research Council (2011).

References

1. Triant DA, Pearson WR (2022) Comparison of detection methods and genome quality when quantifying nuclear mitochondrial insertions in vertebrate genomes. Front Genet 13:984513. https://doi.org/10.3389/fgene.2022.984513

2. Triant DA, DeWoody JA (2007) The occurrence, detection, and avoidance of mitochondrial DNA translocations in mammalian systematics and Phylogeography. J Mammal 88:908–920. https://doi.org/10.1644/06-MAMM-A-204R1.1

3. Triant DA, DeWoody JA (2008) Molecular analyses of mitochondrial pseudogenes within the nuclear genome of arvicoline rodents. Genetica 132(1):21–33

4. Arctander P (1995) Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. Proc R Soc Lond B Biol Sci 262:13–19. https://doi.org/10.1098/rspb.1995.0170

5. Bensasson D (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol 16:314–321. https://doi.org/10.1016/S0169-5347(01)02151-6

6. Hazkani-Covo E, Zeller RM, Martin W (2010) Molecular poltergeists: mitochondrial DNA copies (NUMTs) in Sequenced Nuclear genomes. PLoS Genet 6:e1000834. https://doi.org/10.1371/journal.pgen.1000834

7. Mirol PM, Mascheretti S, Searle JB (2000) Multiple nuclear pseudogenes of mitochondrial *cytochrome b* in *Ctenomys* (Caviomorpha, Rodentia) with either great similarity to or high divergence from the true mitochondrial sequence. Heredity 84:538–547. https://doi.org/10.1046/j.1365-2540.2000.00689.x

8. Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proc Natl Acad Sci 105:13486–13491. https://doi.org/10.1073/pnas.0803076105

9. Triant DA, Dewoody JA (2009) Integrating NUMT pseudogenes into mitochondrial phylogenies: comment on mitochondrial phylogeny of Arvicolinae using comprehensive taxonomic sampling yields new insights. Biol J Linn 97:223–224

10. Wei W, Schon KR, Elgar G et al (2022) Nuclear-embedded mitochondrial DNA sequences in 66,083 human genomes. Nature 611:105–114. https://doi.org/10.1038/s41586-022-05288-7

11. Zhang DX, Hewitt GM (1996b) Nuclear integrations: challenges for mitochondrial DNA markers. Trends Ecol Evol 11:247–251. https://doi.org/10.1016/0169-5347(96)10031-8

12. Alireza M, Morteza N, Reza RH, Mohammad K (2018) Population Genetic structure of transcaucasian mole vole (*Ellobius lutescens*) along Zagros Mountains, Iran. Contemp Probl Ecol 11(2):239–245. https://doi.org/10.1134/S1995425518020075

13. Antunes A, Ramos MJ (2005) Discovery of a large number of previously unrecognized mitochondrial pseudogenes in fish genomes. Genomics 86(6):708–717. https://doi.org/10.1016/j.ygeno.2005.08.002

14. Attardi G, Schatz G (1988) Biogenesis of mitochondria. Annu Rev Cell Biol 4:289–333. https://doi.org/annurev.cb.04.110188.001445

15. Bogdanov AS, Lebedev VS, Zykov AE, Bakloushinskaya IY (2015) Variability of *cytochrome b* Gene and adjacent section of Gene *tRNA-Thr* of mitochondrial DNA in the Northern Mole Vole *Ellobius talpinus* (Mammalia, Rodentia). Genetika 51(12):1433–1438 [in Russian]

16. Calabrese FM, Balacco DL, Preste R, Diroma MA, Forino R, Ventura M, Attimonelli M (2017) NUMTS colonization in mammalian genomes. Sci Rep 7(1):16357. https://doi.org/10.1038/s41598-017-16750-2

17. Cummins JM, Wakayama T, Yanagimachi R (1997) Fate of micro-injected sperm components in the mouse oocyte and embryo. Zygote 5:301–308. https://doi.org/10.1017/S0967199400003889.

18. Dayama G, Emery SB, Kidd JM, Mills RE (2014) The genomic landscape of polymorphic human nuclear mitochondrial insertions. Nucleic Acids Res 42(20):12640–12649. https://doi.org/10.1093/nar/gku1038.

19. Evdokimov NG (2001) Population ecology of the mole-vole. Ekatherinburg [In Russian]

20. Gibb GC, Kennedy M, Penny D (2013) Beyond phylogeny: pelecaniform and ciconiiform birds, and long-term niche stability. Mol Phylogenet Evol 68:229–238. https://doi.org/10.1016/j.ympev.2013.03.021.

21. Hebert PDN, Bock DG, Prosser SWJ (2023) Interrogating 1000 insect genomes for NUMTs: a risk assessment for estimates of species richness. PLoS ONE 18:e0286620. https://doi.org/10.1371/journal.pone.0286620.

22. Kaya A, Coşkun Y (2015) Reproduction, postnatal development, and social behavior of *Ellobius lutescens* Thomas, 1897 (Mammalia: Rodentia) in captivity. Turk J Zool 39:425–431. https://doi.org/10.3906/zoo-1401-73.

23. Kryštufek B, Shenbrot G (2022) Voles and lemmings (Arvicolinae) of the Palaearctic Region. University of Maribor, University Press. https://doi.org/10.18690/um.fnm.2.2022.

24. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096.

25. Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ (1994) NUMT, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J Mol Evol 39:174–190. https://doi.org/10.1007/BF00163806.

26. Lopez JV, Culver M, Stephens JC, Johnson WE, O'Brien SJ (1997) Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. Mol Biol Evol 14:277–286. https://doi.org/10.1093/oxfordjournals.molbev.a025763.

27. Lucas T, Vincent B, Eric P (2022) Translocation of mitochondrial DNA into the nuclear genome blurs phylogeographic and conservation genetic studies in seabirds. R Soc open sci 9:211888. https://doi.org/10.1098/rsos.211888.

28. McArthur K, Whitehead LW, Heddleston JM, Li L, Padman BS, Oorschot V, Geoghegan ND, Chappaz S, Davidson S, San Chin H, Lane RM, Dramicanin M, Saunders TL, Sugiana C, Lessene R, Osellame LD, Chew TL, Dewson G, Lazarou M, Ramm G, Lessene G, Ryan MT, Rogers KL, van Delft MF, Kile BT (2018) BAK/BAX macropores facilitate mitochondrial herniation and

mtDNA efflux during apoptosis. Science 359(6378):eaao6047. https://doi.org/10.1126/science.aao6047

29. Mourier T, Hansen AJ, Willerslev E, Arctander P (2001) The human genome project reveals a continuous transfer of largemitochondrial fragments to the nucleus. Mol Biol Evol 18:1833–1837. https://doi.org/10.1093/oxfordjournals.molbev.a003971.

30. Nacer DF, Do Amaral FR (2017) Striking pseudogenization in avian phylogenetics: NUMTs are large and common in falcons. Mol Phylogenet Evol 115:1–6. https://doi.org/10.1016/j.ympev.2017.07.002.

31. Novikov E, Zadubrovskaya I, Zadubrovskiy P, Titova T (2017) Reproduction, ageing, and longevity in two species of laboratory rodents with different life histories. Biogerontology 18:803–809. https://doi.org/10.1007/s10522-017-9723-7.

32. Parfait B, Rustin P, Munnich A, Rötig A (1998) Coamplification of Nuclear pseudogenes and Assessment of Heteroplasmy of mitochondrial DNA mutations. Biochem Biophys Res Commun 247(1):57–59. https://doi.org/10.1006/bbrc.1998.8666.

33. Pereira RJ, Ruiz-Ruano FJ, Thomas CJE, Pérez-Ruiz M, Jiménez-Bartolomé M, Liu S, Torre J, Bella JL (2021) Mind the *NUMT*: finding informative mitochondrial markers in a giant grasshopper genome. J Zool Syst Evol Res 59:635–645. https://doi.org/10.1111/jzs.12446.

34. Richly E, Leister D (2004) NUMTs in sequenced eukaryotic genomes. Mol Biol Evol 21:1081–1084. https://doi.org/10.1093/molbev/msh110.

35. 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. Syst Biol 61:539–542. https://doi.org/10.1093/sysbio/sys029.

36. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: a laboratory manual, 2nd edn. Cold Spring Harbor, New York

37. Smorkatcheva AV, Kuprina KV (2018) Does sire replacement trigger plural reproduction in matrifilial groups of a singular breeder, *Ellobius tancrei*? Mamm Biol 88:144–150. https://doi.org/10.1016/j.mambio.2017.09.005.

38. Sorenson MD, Fleischer RC (1996) Multiple Independent transpositions of mitochondrial DNA control region sequences to the nucleus. Proc Natl Acad Sci USA 93:15239–15243. https://doi.org/10.1073/pnas.93.26.15239

39. Sorenson MD, Quinn TW (1998) NUMTs: a challenge for Avian Systematics and Population Biology. Auk 115:214–221. https://doi.org/10.2307/4089130

40. Sorenson MD, Ast JC, Dimcheff DE, Yuri T, Mindell DP (1999) Primers for a PCR-Based Approach to mitochondrial genome sequencing in birds and other vertebrates. Mol Phylogenet Evol 12:105–114. https://doi.org/10.1006/mpev.1998.0602

41. Spinks PQ, Shaffer HB (2007) Conservation phylogenetics of the Asian box turtles (Geoemydidae, *Cuora*): mitochondrial introgression, NUMTs, and inferences from multiple nuclear loci. Conserv Genet 8:641–657. https://doi.org/10.1007/s10592-006-9210-1.

42. Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hum Genet 48:370–382

43. Tambovtseva V, Bakloushinskaya I, Matveevsky S, Bogdanov A (2022) Geographic Mosaic of Extensive Genetic Variations in subterranean mole voles *Ellobius alaicus* as a consequence of Habitat Fragmentation and hybridization. Life 12:728. https://doi.org/10.3390/life12050728.

44. Thalmann O, Hebler J, Poinar HN, Pääbo S, Vigilant L (2004) Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes: NUMTS IN APES. Mol Ecol 13:321–335. https://doi.org/10.1046/j.1365-294X.2003.02070.x.

45. van der Kuyl AC, Kuiken CL, Dekker JT, Perizonius WRK, Goudsmit J (1995) Nuclear counterparts of the cytoplasmic mitochondrial 12S rRNA gene: a problem of ancient DNA and molecular phylogenies. J Mol Evol 40:652–657. https://doi.org/10.1007/BF00160513

46. Wallace DC, Stugard C, Murdock D, Schurr T, Brown MD (1997) Ancient mtDNA sequences in the human nuclear genome: a potential source of errors in identifying pathogenic mutations. Proc Natl Acad Sci USA 94:14900–14905. https://doi.org/10.1073/pnas.94.26.14900.

47. Woischnik M, Moraes CT (2002) Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. Genome Res 12(6):885–893. https://doi.org/10.1101/gr.227202

48. Yao YG, Kong QP, Salas A, Bandelt HJ (2008) Pseudomitochondrial genome haunts Disease studies. J Med Genet 45:769–772. https://doi.org/10.1136/jmg.2008.059782.

49. Zhang DX, Hewitt GM (1996a) Highly conserved nuclear copies of the mitochondrial control region in the desert locust *Schistocerca gregaria*: some implications for population studies. Mol Ecol 5:295–300. https://doi.org/10.1046/j.1365-294X.1996.00078.x.

50. Zhang G, Geng D, Guo Q, Liu W, Li S, Gao W, Wang, Yongfei ZM, Wang, Yilin, Bu Y, Niu H (2022) Genomic landscape of mitochondrial DNA insertions in 23 bat genomes: characteristics, loci, phylogeny, and polymorphism. Integr Zool 17:890–903. https://doi.org/10.1111/1749-4877.12582.