Species-specific variation in mitochondrial genome tandem repeat polymorphisms in hares (*Lepus* spp., Lagomorpha, Leporidae) provides insight into their evolution

Riikka Tapanainen[1], Koit Aasumets[1], Zsófia Fekete[1,2], Steffi Goffart[1], Eric Dufour[3¶], & Jaakko L. O. Pohjoismäki[1*¶]

[1]University of Eastern Finland, Department of Environmental and Biological Sciences – Joensuu, Finland

[2]Hungarian University of Agriculture and Life Sciences, Institute of Genetics and Biotechnology, Gödöllő, Hungary

[3]Mitochondrial Bioenergetics and Metabolism, Faculty of Medicine and Health Technology, FI-33014 Tampere University, Finland

*Corresponding author

¶These authors contributed equally

Short title: Species-specific variation of mtDNA tandem repeats in hares

Abstract

The non-coding regions of some vertebrate mitochondrial genomes contain tandem repeat elements, whose copy number can vary between species, populations and even individuals. Both short (~20 bp) and long (130–160 bp) tandem repeats of variable lengths are present in the mitochondrial genomes of hares, rabbits and pika (Lagomorpha), while none can be found

in related mammalian orders. In the presented study, we provide functional annotation for mountain hare (*Lepus timidus*) and brown hare (*L. europaeus*) mitochondrial DNA (mtDNAs) non-coding regions, together with a species as well as population-level analysis of tandem repeat variation. While the mountain hare short tandem repeats (SRs) consist of two randomly repeating but highly conserved 10 bp sequence motifs, brown hares have only one type of 10 bp motif, which shows more sequence variation within the species. The one repeat motif SRs seem unique to brown hares, as the other members of the genus *Lepus* have two motifs. Similarly, the long tandem repeats (LRs) show both sequence and copy number differences between the species. Mountain hares have between four and seven repeats, with median value of five and the brown hares between five and nine, with median value of six repeats. Interestingly, mountain hare mtDNA introgressed into brown hares through hybridization exhibit an intermediate LR copy number distribution. In cultured cells, the transfer of mountain hare mtDNA into mtDNA-less brown hare cells typically result in a gain or loss of one to two LR copies, the new LR haplotype being maintained in heteroplasmy with the old. The reciprocal transfer of brown hare mtDNA had no effect on the LRs. We propose that the tandem repeats are unlikely to have regulatory roles in hare mitochondrial genomes, but rather represent selfish genetic elements, whose maintenance is influenced by the cellular context.

Introduction

Mitochondria are essential cellular organelles best known for production of ATP through process called oxidative phosphorylation (OXPHOS) [1]. They are also a central hub for most metabolic pathways and play specialized functions such as the biosynthesis of iron-sulfur clusters, steroids, fatty acid oxidation, calcium homeostasis, programmed cell death, and more. Due to their origin as a free-living prokaryote, mitochondria have their own small circular DNA genome (mitochondrial DNA, mtDNA), which has undergone reductive evolution associated with gene transfer to the nucleus [2]. In animals, mtDNA is typically less than 20 kb in size

and contains genes for 13 polypeptides of the OXPHOS complexes as well as two mitochondrial ribosomal RNAs and 22 tRNAs required for their translation [3]. The two strands of the mtDNA have very different gene density, with most of the genes being encoded on the so-called heavy strand (H-strand). The designation stems from its high guanine content, giving it a higher mass compared to the complementary light-strand (L-strand), as visualized in alkaline CsCl density gradient centrifugation [4]. Besides the highly compact coding region, the mitochondrial genome also has a non-coding region (NCR), sometimes called the control region due to the presence of control elements required for DNA transcription and replication. The length and organization of the NCR is very variable across the animal kingdom, contributing most to the mtDNA length variation between the species [3].

In vertebrates the mtDNA non-coding region houses curious features, such as the displacement loop (D-loop), which is generated by replication initiating from the main replication origin (OriH: Origin of Heavy strand replication) and terminating prematurely at the "termination associated sequence" (TAS) [5]. As a result, a triple-stranded DNA structure is formed, including the 600–700 nt long nascent H-strand fragment, the 7S DNA, which remains hybridized to its L-strand template. While the human and mouse mitochondrial genomes, probably the best known mtDNAs, show considerable economy in their NCR organization, its size can be variable in other vertebrates due to tandem repeat arrays [6]. Of these, short tandem repeats of around 10 bp are an universal feature for the insect mitochondrial genomes [7], also present in vertebrate species, including mammals [8]. Longer, 100-200 bp repeats are found in certain species of fish [9], birds [10-12] and some mammals, such as rabbits and hares [13]. The repeat number can vary from two [9] to more than one hundred [14] per mtDNA molecule. The presence of mitochondrial tandem repeats is likely overlooked as they can be easily lost during the sequencing and the assembly of short sequence reads, such as the ones obtained from Illumina sequencing platforms.

The tandem repeats on the NCR of the European rabbit (*Oryctolagus cuniculus*) mitochondrial genome were noted already in the 1980s [15] and later work showed that these repeats are widespread also in other mitochondrial genomes of Lagomorpha [13]. There are two types of repeated motifs in rabbits, 20 bp short repeats (SR) and 153 bp long repeats (LR). The SR arrays are expanded or contracted relatively dynamically, with the copy number of the units varying between three to 19. Consequently, the SR arrays can occur in heteroplasmy, i.e. mitochondrial DNAs with different repeat lengths coexisting in the same cell or tissue. An elegant experiment where rabbit SRs were cloned and maintained in a bacterial plasmid demonstrated that slipped-strand mispairing during replication is the main mechanism to explain the dynamic state of the repeat array [16]. The experiments demonstrated that SR insertions are more common than deletions in arrays with less than 10 copies with the opposite being true for longer tracts, resulting the SR lengths to oscillate around an optimal value.

Similar to the SRs, the LRs also have a high mutation rate (10^{-2} per animal per generation), resulting in NCR length variation between generations but also within individuals, manifesting as heteroplasmy as well as mosaicism [13]. In rabbits, LR tend to be present as arrays of five repeats with shorter and longer variants being rarer. While the mtDNA haplotype has no influence of the array length, oddly, some organs such as gonads maintain longer arrays than other tissues [13]. Due to their length, the expansion or deletion of LRs through replication slippage-mispairing is more complicated than for the SRs. In the nucleus, the copy number changes of similar long satellite repeats, including expansions and contractions of rDNA arrays [17], have been proposed to occur through break induced recombination [18]. Interestingly, a similar copy-choice recombination mechanism involving strand invasion of the parental molecule by a free 3′-end of a newly synthetized DNA strand at regions of short sequence homologies has been suggested as a mechanism for mtDNA deletion formation in humans [19,

20]. Non-allelic recombination or end joining could in principle also explain the observed LR length variation.

LR tandem repeats have been found in all Lagomorphs from pikas (*Ochotona* spp.) to hares (*Lepus* spp., *Sylvilagus* spp.) [13]. All LRs have a 20 bp strictly conserved sequence, which has been proposed to be a binding site for a regulatory factor, such as TFAM. In fact, the rabbit LR has been suggested to contain both L-strand (LSP) and H-strand (HSP) promoters for mtDNA transcription [21]. If this is the case, variation in LR length could influence primer synthesis by MTRPOL at these promoters [22], thereby modulating also the replication initiation.

Our group has been interested in hybridization between the mountain hare (*Lepus timidus*) and brown hare (*L. europaeus*) at the northern contact zone of these species in Finland. Despite being separated by three million years of evolution [23], the hybrids are fertile, resulting in gene flow across the species barrier [24, 25]. This gene flow is biased towards the brown hare, which seems to benefit from the hybridization by obtaining locally adapted alleles from the mountain hare [26]. Furthermore, the hybridization results in frequent introgression of mountain hare mtDNA into the brown hare population, encompassing up to fifth of the individuals in certain regions in Finland [25]. When sequencing the mitochondrial genomes of our mountain hare and brown hare cell lines [27], we recently noted a species difference in the length of LRs in the mtDNA non-coding region. A population sampling of 151 mountain hares and 148 brown hares with conspecific mtDNA confirmed that brown hares maintain more LR copies (median = 6) than mountain hares (median = 5), with no correlation with their geographic origin. Interestingly when introgressed into brown hares, the mountain hare mtDNA presented frequent heteroplasmy and gained “brown-hare like” LRs (median = 6 copies). To further test how the nuclear background can influence LRs, we generated cybrids of mountain hare and brown hare fibroblasts with nuclear DNA from one species and mtDNA

from another. The repeat region length changed in three cybrid cell lines out of six, regardless of their genetic background, with the new LR variant maintained in heteroplasmy. In one case, the LR number of brown hare mtDNA was increased from six to seven in cells with mountain hare nucleus, a state which is very rare in the natural population of mountain hares. We conclude that while the nuclear genetic background certainly plays a role in the LR maintenance, environmental factors and differences in cell biology between species also contribute to the observed variation. We also discuss the potential relation of the NCR repeat elements to the regulation of mtDNA.

Materials and Methods

Sampling and DNA isolation

The mountain hare (*Lepus timidus*) and brown hare (*L. europaeus*) specimens presented here are from a larger biobank of 1,202 hare samples collected in 2012–2016 across Finland [24, 25]. The origin and generation of the four mountain hare (LT1, LT4, LT5, LT6) and four brown hare (LE1, LE2, LE3, LE4) fibroblast cell lines has been described elsewhere [27]. Most samples are from hunted animals, with a subset from specimens found dead from the wild. The sampling had minimal impact on the local populations and pose no threat on the habitats. The hunting followed the regional hunting seasons and legislation (Metsästyslaki [Hunting law] 1993/615/5§). The sampling adhered to the ARRIVE guidelines and also no ethical assessment is required. All sampling occurred in Finland and did not involve International Trade in Endangered Species of Wild Fauna and Flora (CITES) or other export of specimens, as defined by the Convention on Biological Diversity (CBD). The DNA was isolated from frozen ear muscle biopsies using the pegGOLD blood & tissue DNA mini kit (VWR) and following the protocol provided by the manufacturer.

Mitochondrial DNA sequencing

The primers used to amplify the hare mtDNA were as follows:

Le93F: TTGTTTTGTAGCAAGTTTACACATGC

Le184R: GCTTAATACCTGCTCCTCTTGATCTA

Le1580F: TTAAACCCATAGTTGGCCTAAAAGC

Le1635R: TTGAGCTTTAACGCTTTCTTAATTGA

Le3045F: AGGCGTATTATTTATCCTAGCAACCT

Le3175R: CCTCATAAGAAATGGTCTGTGCGA

Le3921F: CCCCCTAATCTTTTCCATCATCCTAT

Le4482R: TCATCCTATATGGGCAATTGAGGAAT

Le4689F: AGGCTTTATTCCAAAGTGAATTATTATTCA

Le5417R: AGGCTCCAAATAAAAGGTAGAGAGTT

Le6696F: ATACCGTCTCATCAATAGGCTCCTTC

Le6756R: ATAAAGATTATTACTATTACAGCGGTTAGA

Le8603F: AGCCTATATCTACATGATAATACTTAATGA

Le8698R: CGGATAAGGCCCCGGTAAGTGG

Le10552F: TTGAAGCAACACTAATCCCTACACTA

Le10613R: TCGTTCTGTTTGATTACCTCATCGT

Le11301F: ACCATTAACCTTCTAGGAGAGCTTCT

Le11807R: AGGATAATGATTGAGACGGCTATTGA

Le12407F: GTCTAATCCTAGCTGCTACAGGTAAG

Le12791R: GAGCATAAAAAGAGTATAGCTTTGAA

Le14204F: ATTGTTAACCACTCTCTAATCGACCT

Le14514R: CCAATGTTTCAGGTTTCTAGGTAAGT

Lt16056F: TGGGGTATGCTTGGACTCAAC

Le16119R: TCGTCTACAATAAGTGCACCGG

In total, 12 separate reactions were prepared to cover the mitochondria genome:

1. Lt16056F + Le184R: 1871 bp
2. Le93F + Le1635R: 1543 bp
3. Le1580F + Le3175R: 1596 bp
4. Le3045F + Le4482R: 1438 bp
5. Le3921F + Le5417R: 1497 bp
6. Le4689F + Le6756R: 2068 bp
7. Le6696F + Le8698R: 2003 bp
8. Le8603F + Le10613R: 2011 bp
9. Le10552F + Le11807R: 1256 bp
10. Le11301F + Le12791R: 1491 bp
11. Le12407F + Le14514R: 2108 bp
12. Le14204F + Le16119R: 1916 bp

(Expected fragment size based on the published *Lepus europaeus* mtDNA sequence from Sweden [NC_004028.1]).

The fragments were amplified from total DNA preparations using a PCR program with a 1 min 94 °C denaturing step, followed by 35 cycles of 94 °C for 15 s, 56 °C for 15 s and 72 °C for 2 min and a final 3 min elongation step at 72 °C. The obtained products were gel purified using the GeneJET gel extraction kit (Thermo Fisher Scientific™) and sent for sequencing using Illumina MiniSeq™ at the Genome Center of Eastern Finland.

The sequence of the non-coding region-containing PCR fragment (Lt16056F + Lt184R) was further validated by Sanger sequencing, applying also the following additional sequencing primer:

Le101F: TATAAATTCCTGCCAAACCCCAAAAA

Mitochondrial DNA assembly and annotation

Mitochondrial DNA was assembled using the MitoZ pipeline[28] and the best assembly was selected by comparing the results of the pipeline's outputs with different kmer options. The final assembly was done using the un options --clade Chordata –fastq_read_length 150, --requiring_taxa Chordata --genetic_code 2 --kmers_megahit 21 29 39 59 79 99 119 141. The pipeline included the tools fastp [29] for cleaning the raw data, MEGAHIT [30] for assembly, after which sequences were filtered using HMMER [31] to ensure the correct taxa and the completeness of protein-coding genes. Annotation was performed using TBLASTN [32], GeneWise [33] and MiTFi [34]. The annotation of the non-coding region (NCR) as well as the illustration of the mitochondrial genome was done with Geneious® 10.2.6 (Biomatters. Available from https://www.geneious.com). The functional loci on the NCR were identified based on the similarity with the human (NC_012920) and mouse (FJ374652) NCR sequences.

Genotyping of the long repeat (LR) region

The LR length genotyping of the hare samples was performed by PCR using the following primers that bind the flanking regions of the repeat run (Fig. 1A, B):

Lepus timidus mtDNA

LtLR-F: AGAACCGTGACATAGCACTTACTTTC

LtLR-R: TAACATATTGGTGTTAGAATGTTTTTAGT

L. europaeus mtDNA

222 LeLR-F: TATAAATTCCTGCCAAACCCCAAAAA

223 LeLR-R: GCTTAATACCTGCTCCTCTTGATCTA

224

225 The used PCR program had an initial denaturation step of 94°C for 2 min, followed by 35
226 cycles of 94°C for 20 s., 59°C for 20 s. and 72°C for 1:30 min, with final elongation at 72°C
227 for 5:00 min, using AccuStart II™ PCR SuperMix (Quantabio). The PCR products were
228 separated over 1 % TAE agarose gel. In general, the Le primer pair works also for mountain
229 hares, but did not reliably amplify certain population samples.

230

231 **Analyses of the SRs**

232 Core SR sequences and their flanking regions of the Finnish *Lepus timidus* Finland and *Lepus*
233 *europaeus* were manually retrieved from the whole mitochondrial genome sequences. All
234 possible core dimers (1 for LE type, 4 for LT) were then used as query sequence for a basic
235 nucleotide BLAST search. Results were filtered to only include mitochondrial genomes
236 sequences from defined geographic isolates (LE query) or from one genome per *Lepus* species
237 (LT queries). For LT queries; *Lepus tibetanus*: MN539746.1, *L. tolai*: MN539744.1, *L. arcticus*:
238 NC_044769.1, *L. sinensis*: NC_025316.1, *L. coreanus*: NC_024259.1, *L. granatensis*:
239 NC_024042.1, *L. othus*: KJ397608.1, *L. corsicanus*: KJ397606.1, *L. capensis*: NC_015841.1
240 and *L. yarkandensis*: MN539747.1 mitochondrial genomes were selected. *L. townsendii* was
241 excluded due to poor quality sequencing of its SR region. For the LE query, *L. europaeus*
242 Poland 3: KY211034.1, Poland 2: KY211033.1, Poland 1: KY211032.1, Germany 2:
243 KY211031.1, Germany 1: KY211030.1, Greece 4: KY211029.1, Greece 3: KY211028.1,
244 Greece 1: KY211026.1, Cyprus 4: KY211025.1, Cyprus 3: KY211024.1, Cyprus 2:
245 KY211023.1, Cyprus 1: KY211022.1, and Turkey 1: KY211021.1 were selected. LE Greece 2
246 isolate was excluded as is SR region was identical to that of Greece 3 isolate. Rabbit

247 (*Oryctolagus cuniculus*) reference genome mtDNA (NC_001913.1) was used for a comparison.
248 The proportion of mutated core repeats was compared using Fisher exact test. Because of its
249 unique SR features, *L. yarkandensis* was excluded from the statistical analyses. Sequence
250 uncertainties in the repeats of *L. otus* (1) and *L. corsicanus* (3) were all considered as mutated
251 for the Fisher exact test, the mutation frequency was nevertheless significantly higher in LE-
252 type of SRs. To visualize the frequency of mutation in function of the repeat location, the repeat
253 positions were normalized to the number of repeats in the SR leading to an x-axis ranging from
254 0 ("beginning" of the SR region) to 1 ("end" of the SR region).
255

256 **Generation of the cybrid cell lines**

257 Cybrid cells were generated using the chemical enucleation method [35]. In brief, we first
258 generated recipient mtDNA-less ρ0 cells from immortalized LT1, LT4, LT6, LE1, LE2 and LE3
259 fibroblast cell lines by culturing them in the presence of 100 µM ddC, until mtDNA was not
260 detected anymore on a Southern blot. The nucleus of the mtDNA donor cells was killed with
261 Actinomycin D, with the effective concentration determined separately for each cell line. The
262 optimal concentration for most cell lines was 1 µg/ml for 18h, except for LT1 (1 µg/ml for 15h)
263 and LE2 (2 µg/ml for 28h). The cybrid fusions were conducted in both directions (same cell
264 line either a recipient or a donor of mtDNA) for the following cell line pairs: LE1♂ x LT4♀,
265 LE2♀ x LT1♂ and LE3♀ x LT6♂. For the control cybrids with mtDNA from the same species
266 but a different cell line than the nucleus, we used LE1♂(nucleus) x LE2♀(mtDNA) and LT4♀
267 (mtDNA) x LT6♂ (nucleus) combinations. The fusion was performed by growing the mtDNA
268 donor cells on 6-well plates and treating them with Actinomycin D. After replacing the
269 treatment medium with fresh DMEM (10% FBS) containing 1 million ρ0 cells were added to
270 the well for 3h. The cells were then washed 3 × with fresh DMEM before addition of 45%
271 polyethylene glycol (PEG, MW: 1450) for 60s. The cells were then washed again 3 × with

DMEM + 10% DMSO and 1 × with DMEM. Finally, the cells were given DMEM + 10% FBS + P/S. The cybrids were grown until they reached confluency, at which stage half of the cells in a well were collected for genotyping. DNA was extracted with Quick-DNA Miniprep Kit (Zymo Research).

As the donor and recipient cells differed by their sex, PCR-RFLP of *ZFX* and *ZFY* loci [36] was used to genotype the cybrid nucleus and *CYTB* PCR-RFLP their mtDNA [37]. Each original cybrid pool was genotyped and promising looking pools were subcloned on 96-wells. These clonal cell lines were rechecked for the expected nuclear and mtDNA haplotypes, obtaining the final cybrid cell lines. The ρ0 and cybrid cells were maintained in high glucose (4.5 g/l) DMEM, supplemented with 10% FBS and 50 µg/ml uridine. The cybrid cells were otherwise grown the same, but without added uridin to select for functional mitochondrial DNA.

Agarose gel electrophoresis and Southern hybridization

Mitochondrial DNA from LE and LT parental and cybrid cell lines was isolated using phenol:chloroform extraction and ethanol precipitation. 2 µg of uncut mitochondrial DNA was separated over a 0.4 % agarose gel in 1x TBE buffer at 25 V for overnight at room temperature. To facilitate the transfer of mtDNA onto nylon membrane by capillary transfer the gel was subjected to acid depurination followed by denaturation to create suitable hybridization targets. Transferred DNA was fixed to the blot by baking 80 °C for 2 h. Southern blot was probed using (α-32P)-dCTP labelled PCR probe spanning hare mtDNA nucleotides 16840-88 or 17077-88 at 65 °C for overnight. Signals were captured on a phosphor screen and detected using phosphorimager (GE Healthcare).

296 ### Statistical tests

297 The RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) for
298 predicting secondary structures of single stranded RNA or DNA sequences was used to screen
299 for hairpin formation in the mitochondrial non-coding region and the QGRS mapper [38] to
300 identify putative Quadruplex forming G-Rich Sequences.

301 To test whether there was a geographical pattern in the LR length variation among
302 Finnish hares, a spatial autocorrelation analysis was used applying the Moran's I test with sf
303 and spdep packages in R (https://CRAN.R-project.org/).

304 ## Results

305 ### The structure of the mitochondrial DNA non-coding region in hares

306 We amplified the mitochondrial genomes from four mountain hare (LT1, LT4, LT5, LT6) and
307 four brown hare (LE1, LE2, LE3, LE4) cell lines as 12 overlapping PCR products, which were
308 subsequently sequenced using Illumina Miniseq™. When assembling the sequences, we noted
309 that the obtained sequence for the NCR was shorter than expected from the PCR product
310 lengths. Closer inspection revealed that this was because the short sequence reads (2 × 150 bp)
311 over the LR region were assembled as a single repeat element due to their redundancy. The
312 PCR products containing NCR elements were therefore reanalyzed using Sanger sequenced to
313 clarify LR copy number and finalized the assemblies (Table 1, Fig. 1). Interestingly, our
314 mountain hare as well as brown hare mitochondrial genomes are smaller (Table 1) than the
315 previously published mountain hare mtDNA from a Finnish specimen (NC_024040.1) (17,755
316 bp) or the brown hare mtDNA from Sweden (NC_004028.1). The size difference is explained
317 by the length of the non-coding region (NCR), where our LT cell lines have five and LE cell
318 lines six long tandem repeats (LRs) compared to seven in both previous mitochondrial
319 genomes.

320 The length of the NCR is 2,038 bp in LT1 (Fig. 2A) and 2,102 bp in LE1 (Fig. 2B). It
321 contains recognizable features such as a rather conserved region between the termination
322 associated sequence (TAS) and the cluster of short tandem repeats (SR), which by length and
323 location could correspond the D-loop region containing the 7S DNA. The SRs and the LRs are
324 separated by a variably long sequence with conserved sequence blocks (CSBs) 2 and 3, thought
325 to be relevant for the control of mtDNA gene expression and replication priming [39].
326 However, if replication is primed at or near the CSBs, the primer 3´-ends are relatively far away
327 from the known 5´-ends of DNA, considered to correspond the replication origin and start of
328 the 7S DNA [40]. In fact, Southern blot analysis of this region can only detect the 7S DNA
329 when using a probe that extends downstream of the SRs (Fig. 2C). Also, the size of the 7S does
330 not differ between the cell lines despites the difference in the length of the NCR-region. Of
331 note, we could confirm that Lagomorphs do not seem to have any CSB1[13, 21]. The CSB2
332 differs by a couple of indels between the species, the CSB3 seems highly conserved (Fig. 3A).
333

334 **Hare SR elements consist of 10 bp core motifs**

335 At first glance, the mountain hare (LT) SRs appear as a 20 bp repeat element similar in length
336 to the rabbits SRs (Fig. 3A). However, these are not consistent, and part of the SR arrays
337 showed repetition of only half of the sequence. In closer inspection, the element consists of
338 two 10 bp motifs, TACGCGCACG, and TACACCCACG, which differ by the two underlined
339 nucleotides. In contrast, the brown hare (LE) SRs contain only one type of 10 bp core motif
340 (TACACGCACG), which differs from both LT motifs by one of the variable nucleotides (Fig.
341 3B). Interestingly, the 20 bp rabbit SR motif contains the 7 first nucleotides of the brown hare
342 (LE) SR motif spliced together with the last 8 nucleotides of LT SR motif1, separated by an
343 unique 5 bp (GTACG) spacer sequence (Fig. 3B).

While the LE-type is present only in brown hares, the two LT-types occur in all other hare species that we analyzed (Table 2), with the two core elements alternating variably in the SR arrays. Curiously, also the cape hare (*Lepus capensis*), a close relative of brown hare [23], has LT-type SRs. *Lepus yarkandensis* constitute another curiosity, with randomly alternating core 1 and 2 LT-type SR motifs being separated by di-nucleotides (TT or TC). Overall, the LT-type SRs show considerably less variation than the LE-types (Fig. 3C, 3.35%; $p < 0.001$, Fisher's exact test), despite being spread across many species. The copy number of the LT core repeat sequences varied from 32 in our LT cell lines and other *L. timidus, coreanus, granatensis, otus* and *corsicanus* isolates to 16 in *L. tibetanus* (Table 2). The copy number of the LE type, present only in brown hares, ranged from 28 to 32.

As repetitive sequences are likely to form secondary structures, we were interested to see whether the SR sequences of the two hares can form hairpins when single-stranded. This is particularly interesting as hairpin formation at the second most prominent replication origin on mitochondria, the light-strand origin (OriL) located at the WANCY tRNA cluster, is proposed to be required for L-strand replication priming [41, 42]. A rather strong hairpin was predicted for the H-strand SRs from the mountain hares but not from the brown hares, nor on the L-strand SR sequences of either species (Fig. 4A–C). As for other structural features, the entire non-coding region contained only a single sequence stretch at the 5′-end of the 7S DNA predicted to form a strong guanine quadruplex (G4) (Fig. 4D).

LR length variation in Finnish hares

The LRs of the two species are similar in length (Fig. 5A), 139 bp in the mountain hare and 140 bp in the brown hare, differing at 21 nucleotide locations and containing the same 20 bp conserved sequence element as previously reported for all Lagomorpha [13]. As also noted earlier, the LR copy number shows variation between and even within individuals (Fig. 5B, C).

369 Although PCR tends to cause the amplification of short repeat sequences as an artefact, which
370 are evident as a faint ladder of PCR products of varying repeat content, the common repeat
371 haplotype was readily detectable as a strong band in all population samples and systematically
372 confirmed the sequencing results.

373 While the mechanism of dynamic length variation in the SRs has been shown to result
374 from replication slippage [16], the factors influencing the LR array length are less clear. As
375 they are also known to have species differences [6], we were interested to see whether the LR
376 array length would show population-specific variation. The most common length of LR was
377 five repeats in Finnish mountain hares, with six repeats being the next abundant (Fig. 6A). The
378 shortest observed repeat was four and longest seven, which was represented only by one
379 sample. Curiously, the only previously published mountain hare mtDNA from Finland
380 (NC_024040.1) also has seven repeats. Heteroplasmy was detected in five samples (3 %). No
381 geographic pattern in the LR length variation was observed among the 151 mountain hares
382 across Finland (Moran I statistic standard deviate = 8.09×10^{-10}, p-value = 0.5).

383 In contrast to mountain hares, the most common LR-region length in brown hares was
384 six repeats, with seven and eight repeats being more common than five (Fig. 6A). The longest
385 observed LR-region had nine repeats, while shorter than five repeats occurred only in
386 heteroplasmic samples (Fig. 5C). In the heteroplasmic samples, the shortest repeat sequence
387 length was three copies. Overall, brown hares showed more frequent heteroplasmy (12 %, p =
388 0.0045, Fisher exact test). Interestingly, brown hares with introgressed mountain hare mtDNA
389 showed an intermediate LR-region length distribution mtDNA (Fig. 6A). Although they had
390 the same median length of six repeats as brown hares with conspecific mtDNA, five repeats
391 were much more common (21% vs. 2%). Also a notable proportion (19 %) of the introgressed
392 samples were heteroplasmic, although the difference to the other brown hares was not
393 significant (p = 0.3077, Fisher exact test). Again, no correlation between the LR length and

394 geography was observed in brown hares with or without introgressed mtDNA. Only a few
395 mountain hares with brown hare mtDNA were present in our collection [25] and therefore were
396 not included in the analysis.
397

398 ### Transfer of mtDNA between cells causes instability in the LR copy number

399 Our LT and LE fibroblasts had respectively five and six LR repeats. If the repeat length is an
400 autonomous property of mitochondria, we expect it to remain stable despite the cohabitation
401 with a different nuclear genome. If it is controlled by the nuclear genome we expect it to be
402 rapidly converted to the copy number corresponding to the nuclear donor. While no change in
403 the LR copy number was observed when mtDNA was transferred from mountain hare to brown
404 hare cells (Fig. 6B), the six-copies LR array length of brown hare mtDNA increased to seven
405 in LT4, remained the same in LT1 and decreased to five in LT6 nuclear background (Fig. 6C).
406 Curiously, novel LR copy numbers of four and three appeared also in control cybrids with LT4
407 mtDNA in LT6 nuclear background (Fig. 6B), with no change observed in the brown hare
408 control cybrids (Fig. 6C). In all cases the new LR array haplotypes existed in heteroplasmy
409 with the original haplotype.

410 ## Discussion

411 When sequencing the mitochondrial DNAs as a part of the basic characterization of our hare
412 fibroblast cell lines [27], we noted a discrepancy between the expected genome size and the
413 assemblies obtained from short-read Illumina data. Closer inspection revealed that this was
414 because of misassembly of the part of the non-coding region with known long tandem repeats
415 (LRs). We corrected the assemblies with Sanger sequencing and provide here the fully
416 annotated mitochondrial genomes for four mountain hares and four brown hares from Finland
417 (Table 1, Figs. 1–2). Although tandem repeats are a known elements of the non-coding region
418 in many vertebrate species, including lagomorphs, they remain an overlooked aspects of

419 mitochondrial genome variation, largely due to their absence in humans and other well-studied
420 model organisms. In addition, and as in our case initially, they are likely to be lost when
421 assembling mitochondrial genome sequencing data obtained using short-read sequencing
422 platforms. For this reason, a lot of hare mitochondrial genome assemblies available in GenBank
423 have only one LR-element and many of the SR-array sequences might have issues as well. It
424 is likely that the same problem applies to the mitochondrial genome assemblies from other
425 organisms, meaning that the frequency and diversity of such elements in vertebrates is probably
426 underestimated. Recent development of long-molecule sequencing technologies can help to
427 complete our understanding of the repeat element variation, especially as the full-length
428 mitochondrial genomes can be assembled reliably from whole-genome sequencing data [43].

429 Not much is known about the function and biological significance of the repeat
430 elements in the non-coding region of mitochondrial genomes. Initial studies proposed that the
431 Lagomorph LRs are involved in the regulation of transcription and replication initiation [13,
432 21]. However, there are other conserved regulatory elements on the NCR (Fig. 2), which are
433 known to function in e.g. mtDNA replication priming in other mammals [39]. Furthermore, the
434 7S DNA, resulting from prematurely terminated replication, does not contain sequences
435 between LR- and SR-regions (Fig. 2C), indicating that the replication is likely to start at the
436 SR, as suggested by some authors [44].

437 Interestingly, the SR array length can change rather dynamically in rabbits [13, 16, 21]
438 and their copy number is highly variable (from three to 19). However, in hares the nature of
439 the repeating sequences made the delimitation of the actual repeat element difficult, as different
440 types of varying lengths repeats can be identified (Fig. 3A). Like in rabbits, a 20 bp SR element
441 can be identified in mountain hares. However, this element consists in fact of two separate
442 motifs (Fig. 3B), which can occur alone as the repeating element in the SR array. In brown
443 hares the SR arrays consist only of one core 10 bp motif, which differs from both mountain

hare motif types by a single nucleotide (Fig. 3B). If only these canonical sequences are counted, their copy number in the different hare species varied between 16 and 32 (Table 2). Considering the evolution of the SRs, it is interesting that the LT and LE core motif ends for also the majority of the 20 bp rabbit, interrupted by five differing nucleotides GTACG. It is noteworthy that these can be converted by C>T and A>G transitions to form a LT motif 2 + LE motif 20 bp consensus sequence block (Fig. 3B). It is likely that such transition mutations can occur repeatedly during evolution, while the G>C transversion, separating the LT motif 2 and 3´part of the rabbit motif from the rest (Fig. 3B), is unlikely as a recurrent point mutation. Alternatively, it is possible that the LE motif and LT motif 2 are recombinants of two ancestral repeat types, which are present (in modified state) in rabbits. Overall, the remarkable conservation of the core LT-type SR repeats in hares (*Lepus*) is intriguing (Fig. 3C). This suggests that their origin predates the separation of these species and highlights the unique situation of *L. europaeus* SRs, whose core sequence is not shared by the other *Lepus* species, including its close relative the cape hare (*L. capensis*) [23]. Nevertheless, the identical length of the SRs core sequences (10 bp) and their similarity in all *Lepus* species (Fig. 3B, only a single nucleotide difference *separates L. europaeus* core sequence from each of two SR cores found in other *Lepus* species) can be only explained by common ancestry or mechanism of origin among all species in the genus. The sequence homogeny of the SR runs within a species is quite likely maintained by constant contraction and reamplification of the repeat motifs, resulting in their rapid drift.

Three more observations are worth mentioning concerning the SRs region. (a) In all species and independently of the SR core, the repeats located at the 3´ end and in the center of the SR are the least conserved (Fig. 3C). Similar asymmetry is known to result from the unidirectional gene conversion associated with some mitochondrial mobile elements or with mating type switching in yeast [45, 46], which includes a double-strand break followed by erosion of the ends that generates nucleotide variation after repair. (b) Amongst the 48 core

469 sequence variants in LE-type and LT-type of repeats, and even though the SR core sequences
470 show very strong homology, no LT-type core variant could be observed in species with LE-
471 type core repeats and *vice versa*. This is curious, as it suggests that a unique SR motif
472 amplification event has taken place specifically in the brown hare evolutionary lineage.
473 Sequencing mitochondrial SRs from more *Lepus* species and isolates would be needed to allow
474 drawing further hypotheses from this observation. (c) One species *L. yarkandensis* presents a
475 unique SR motive derived constituted of LT motif 1 + LT motif 2 + a two-nucleotide (TT or
476 TC) spacer. Similar short footprint signatures have been associated to the excision of some
477 transposons [47] as well as non-homologous end joining repair mechanisms [48].

478 The biological significance or functional roles of the SRs remain enigmatic. It is plausible
479 that the SRs could form secondary structures when being single-stranded, which could promote
480 priming by MTRPOL, as occurs during the initiation of L-strand replication at OriL [42].
481 Interestingly, hairpin formation is predicted to be stronger for the H-strand (Fig. 4), which –
482 following the OriL priming mechanism – would indicate a L-strand origin. Similarly, the
483 reported variation in the SR copy number [13, 21] and their proneness to mutate through strand-
484 slippage [16], are counterintuitive features for important genomic functions such as the
485 regulatory control of DNA-replication.

486 Like the short repeats, the long repeats contain a core (20 bp) sequence segment (Fig.
487 4A), which is conserved in all lagomorphs [13, 21]. Coincidentally, this is the same length as
488 the duplex SR motifs, although the sequence bears no resemblance to them and has no obvious
489 sequence features, such as palindromes. If it represents a regulatory element capable of binding
490 protein factors such as TFAM, as proposed [21], the existence of multiple such sites on mtDNA
491 is somewhat reminiscent of the initiator titration model of bacterial replication initiation [49,
492 50]. In *E. coli* for example, the initiation factor DnaA is titrated away from the replication origin
493 *oriC* binding sites by providing dispersed competing recognition sites dispersed elsewhere on

494 the genome. Replication of the genome causes the duplication of these sites, binding more
495 DnaA and thereby reducing its availability to initiate new rounds of replication. The obvious
496 problem for this mechanism to operate in the regulation of hare mtDNA replication initiation
497 is the variability in the copy number of the LRs (Fig. 4B, C), where having three to nine copies
498 of competitive binding sites per genome would effectively negate any accurate regulation,
499 although this could be balanced by difference in mtDNA copy number and TFAM molecules.
500 Unlike the dispersed DnaA sites, the LRs are clustered together and situated so that their
501 doubling would occur only at the end of the replication. Thus, they are unlikely to prevent re-
502 initiation before termination. Still, they could control replication by titrating TFAM in function
503 of the number of mtDNA copies (and therefore SRs repeats) already present in the
504 mitochondria. However, as with the SRs, if these elements have such an important role in
505 genome regulation, why would they be present in some evolutionary lineages but not in others?
506 For example, rodents (Rodentia) – a sister order of Lagomorpha – as well as most other
507 mammals do not have such repetitive sequence arrays on their mtDNAs.

508 The species difference in the LR length distribution between brown hare and mountain
509 hare (Fig. 6A) is interesting, as it tells that their variation between individuals is not random.
510 In fact, similar differences in mtDNA tandem repeat lengths are known from other closely
511 related species [6, 10, 11, 51]. As the species difference is fundamentally a genetic difference,
512 one might expect that the LR copy number would then also be governed by the nuclear genome.
513 However, the fact that LR copy number can vary between the tissues of the same individual in
514 rabbits and that certain tissues are more prone to have longer arrays than others [13], suggests
515 a more complicated mechanism, where the nuclear genetic background is only one variable.
516 As different tissues have different metabolic requirements, manifesting as differences in
517 mitochondrial activity and mtDNA maintenance strategies [52], it is plausible that the internal
518 environment in mitochondria also indirectly effects the LR-region length through factors such

as temperature, oxidative stress and replication/repair protein availability. In fact, we have previously noted that the mountain hare fibroblasts maintain higher mitochondrial membrane potential than the brown hare cells [27], as well as have other differences in their oxidative metabolism. In this case, the energetic state of the cells can be influenced by variables such as metabolic adaptations, tissue type, ageing and stress, causing also the observed variation in the LR copy numbers. If the LRs are generated or lost as a result of strand-slippage during replication, as seems to be the case with the SRs [16], changes in the mitochondrial internal environment could increase the probability of such events, resulting in LR copy number change. Similarly, factors such as mitochondrial temperature [53], could favor certain repeat lengths over others and stabilize the LR copy number variation. Overall, it is curious that both the SR core sequence shows less conservation (Fig. 3C) and the LR arrays show more length variation in brown hares than in mountain hares.

Interestingly, our experiments with the cybrids showed mixed results (Fig. 6B). First, the effect on LR copy number was asymmetric and only presented by cells with mountain hare nucleus, regardless of the mtDNA origin. Also, the change in the LR array length in was not consistent. While in one cybrid line the brown hare mtDNA lost a LR copy, obtaining the common copy number of five for a mountain hare mtDNA, in another the copy number was increased to seven, a LR array length not presented by any of the parental cell lines. In contrast, shorter than parental mountain hare LR arrays appeared in the control cybrid with LT6 nucleus and LT4 mtDNA. Although the sample size is small, it is interesting that the observed effect is consistent with the frequent introgression of mountain hare mtDNA into brown hare population but not vice versa [25, 54], suggesting that mtDNA maintenance is perhaps more permissive or resilient in cells with brown hare nucleus. Alternatively, the mountain hare cells could be more sensitive to the stress caused by the process of generating cybrids. In all cases, the new

haplotype appeared in heteroplasmy with the old, showing that they arise as a spontaneous rearrangement of the existing LRs and increase in abundancy through drift or selection.

As a conclusion, although being a decades old discovery, repeat elements on the non-coding region remain an overlooked feature of Lagomorph mitochondrial genomes. We recognized conserved 10 bp core sequences in the short repeats, with two different motifs present in most hare species and one in brown hares. Interestingly, SR core element copies showed sequence variation in brown hares, but not in other hare species. We also found that while the long repeat arrays are dynamic, their presentation is not random, as the different species show different length distributions. Considering their absence from most mammalian species, as well as the lack of clear biological role in mtDNA maintenance and expression, it is likely that both repeat types have risen and are maintained as selfish genetic elements. It is unlikely that these elements are mobile and they are certainly too small to encode functional gene products, their reproduction has to be related to mtDNA replication mechanisms. Why such elements have not been lost from otherwise compact mitochondrial genomes and how species-specific copy number variation is maintained, warrants for further research.

Acknowledgements

Dr Seppo Helisalmi and Dr Joose Raivo from the Gene diagnostics lab at Genome Center of Eastern Finland are thanked for providing the mtDNA sequencing service. We thank Ms Anita Kervinen for her valuable laboratory assistance in preparing the mtDNA PCRs for the sequencing. This research is produced by the xHARES consortium, funded by the Academy of Finland (grant no. 329264 for ED and JLOP).

References

1. Spinelli JB, Haigis MC. The multifaceted contributions of mitochondria to cellular metabolism. Nat Cell Biol. 2018;20(7):745-54. Epub 2018/06/29. doi: 10.1038/s41556-018-0124-1. PubMed PMID: 29950572; PubMed Central PMCID: PMCPMC6541229.
2. Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol. 2003;29(3):380-95. Epub 2003/11/15. doi: 10.1016/s1055-7903(03)00194-5. PubMed PMID: 14615181.
3. Boore JL. Animal mitochondrial genomes. Nucleic Acids Res. 1999;27(8):1767-80. Epub 1999/04/02. doi: 10.1093/nar/27.8.1767. PubMed PMID: 10101183; PubMed Central PMCID: PMCPMC148383.
4. Kasamatsu H, Robberson DL, Vinograd J. A novel closed-circular mitochondrial DNA with properties of a replicating intermediate. Proc Natl Acad Sci U S A. 1971;68(9):2252-7. Epub 1971/09/01. doi: 10.1073/pnas.68.9.2252. PubMed PMID: 5289384; PubMed Central PMCID: PMCPMC389395.
5. Jemt E, Persson O, Shi Y, Mehmedovic M, Uhler JP, Davila Lopez M, et al. Regulation of DNA replication at the end of the mitochondrial D-loop involves the helicase TWINKLE and a conserved sequence element. Nucleic Acids Res. 2015;43(19):9262-75. Epub 2015/08/09. doi: 10.1093/nar/gkv804. PubMed PMID: 26253742; PubMed Central PMCID: PMCPMC4627069.
6. Lunt DH, Whipple LE, Hyman BC. Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. Mol Ecol. 1998;7(11):1441-55. Epub 1998/11/20. doi: 10.1046/j.1365-294x.1998.00495.x. PubMed PMID: 9819900.
7. Solignac M, Monnerot M, Mounolou JC. Concerted Evolution of Sequence Repeats in Drosophila Mitochondrial-DNA. J Mol Evol. 1986;24(1-2):53-60. doi: Doi 10.1007/Bf02099951. PubMed PMID: WOS:A1986F784400006.

8. Savolainen P, Arvestad L, Lundeberg J. mtDNA tandem repeats in domestic dogs and wolves: mutation mechanism studied by analysis of the sequence of imperfect repeats. Mol Biol Evol. 2000;17(4):474-88. Epub 2000/03/31. doi: 10.1093/oxfordjournals.molbev.a026328. PubMed PMID: 10742040.

9. Wang TY, Tzeng CS, Teng HY, Chang T. Phylogeography and identification of a 187-bp-long duplication within the mitochondrial control region of (Teleostei: Balitoridae). Zool Stud. 2007;46(5):569-82. PubMed PMID: WOS:000250416200004.

10. Mundy NI, Helbig AJ. Origin and evolution of tandem repeats in the mitochondrial DNA control region of shrikes (Lanius spp.). J Mol Evol. 2004;59(2):250-7. Epub 2004/10/16. doi: 10.1007/s00239-004-2619-6. PubMed PMID: 15486698.

11. Omote K, Nishida C, Dick MH, Masuda R. Limited phylogenetic distribution of a long tandem-repeat cluster in the mitochondrial control region in Bubo (Aves, Strigidae) and cluster variation in Blakiston's fish owl (Bubo blakistoni). Mol Phylogenet Evol. 2013;66(3):889-97. Epub 2012/12/06. doi: 10.1016/j.ympev.2012.11.015. PubMed PMID: 23211719.

12. Wang X, Liu N, Zhang H, Yang XJ, Huang Y, Lei F. Extreme variation in patterns of tandem repeats in mitochondrial control region of yellow-browed tits (Sylviparus modestus, Paridae). Sci Rep. 2015;5:13227. Epub 2015/08/20. doi: 10.1038/srep13227. PubMed PMID: 26288099; PubMed Central PMCID: PMCPMC4541255.

13. Casane D, Dennebouy N, de Rochambeau H, Mounolou JC, Monnerot M. Nonneutral evolution of tandem repeats in the mitochondrial DNA control region of lagomorphs. Mol Biol Evol. 1997;14(8):779-89. Epub 1997/08/01. doi: 10.1093/oxfordjournals.molbev.a025818. PubMed PMID: 9254915.

14. Hoelzel AR. Evolution by DNA turnover in the control region of vertebrate mitochondrial DNA. Current Opinions in Genetics and Development. 1993;3:891-5.

15. Ennafaa H, Monnerot M, Elgaaied A, Mounolou JC. Rabbit Mitochondrial-DNA - Preliminary Comparison between Some Domestic and Wild Animals. Genet Sel Evol. 1987;19(3):279-88. doi: DOI 10.1051/gse:19870301. PubMed PMID: WOS:A1987K245500001.

16. Pfeuty A, Gueride M, Lecellier G. Expansion/contraction of mammalian mitochondrial DNA repeats in Escherichia coli mimics the mitochondrial heteroplasmy. J Mol Biol. 2001;314(4):709-16. Epub 2001/12/06. doi: 10.1006/jmbi.2001.5160. PubMed PMID: 11733991.

17. Kobayashi T, Ganley AR. Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. Science. 2005;309(5740):1581-4. Epub 2005/09/06. doi: 10.1126/science.1116102. PubMed PMID: 16141077.

18. Thakur J, Packiaraj J, Henikoff S. Sequence, Chromatin and Evolution of Satellite DNA. Int J Mol Sci. 2021;22(9). Epub 2021/05/01. doi: 10.3390/ijms22094309. PubMed PMID: 33919233; PubMed Central PMCID: PMCPMC8122249.

19. Phillips AF, Millet AR, Tigano M, Dubois SM, Crimmins H, Babin L, et al. Single-Molecule Analysis of mtDNA Replication Uncovers the Basis of the Common Deletion. Mol Cell. 2017;65(3):527-38 e6. Epub 2017/01/24. doi: 10.1016/j.molcel.2016.12.014. PubMed PMID: 28111015.

20. Persson O, Muthukumar Y, Basu S, Jenninger L, Uhler JP, Berglund AK, et al. Copy-choice recombination during mitochondrial L-strand synthesis causes DNA deletions. Nat Commun. 2019;10(1):759. Epub 2019/02/17. doi: 10.1038/s41467-019-08673-5. PubMed PMID: 30770810; PubMed Central PMCID: PMCPMC6377680.

21. Dufresne C, Mignotte F, Gueride M. The presence of tandem repeats and the initiation of replication in rabbit mitochondrial DNA. Eur J Biochem. 1996;235(3):593-600. doi: DOI 10.1111/j.1432-1033.1996.00593.x. PubMed PMID: WOS:A1996TU39700019.

22. Kuhl I, Miranda M, Posse V, Milenkovic D, Mourier A, Siira SJ, et al. POLRMT regulates the switch between replication primer formation and gene expression of mammalian mtDNA. Sci Adv. 2016;2(8):e1600963. Epub 2016/08/18. doi: 10.1126/sciadv.1600963. PubMed PMID: 27532055; PubMed Central PMCID: PMCPMC4975551.

23. Ferreira MS, Jones MR, Callahan CM, Farelo L, Tolesa Z, Suchentrunk F, et al. The Legacy of Recurrent Introgression during the Radiation of Hares. Syst Biol. 2021;70(3):593-607. Epub 2020/12/03. doi: 10.1093/sysbio/syaa088. PubMed PMID: 33263746; PubMed Central PMCID: PMCPMC8048390.

24. Levanen R, Thulin CG, Spong G, Pohjoismaki JLO. Widespread introgression of mountain hare genes into Fennoscandian brown hare populations. Plos One. 2018;13(1). doi: ARTN e0191790

10.1371/journal.pone.0191790. PubMed PMID: WOS:000423416600098.

25. Levanen R, Kunnasranta M, Pohjoismaki J. Mitochondrial DNA introgression at the northern edge of the brown hare (Lepus europaeus) range. Annales Zoologici Fennici. 2018;55(1-3):15-24. PubMed PMID: WOS:000431406800002.

26. Pohjoismaki JLO, Michell C, Levanen R, Smith S. Hybridization with mountain hares increases the functional allelic repertoire in brown hares. Sci Rep. 2021;11(1):15771. Epub 2021/08/06. doi: 10.1038/s41598-021-95357-0. PubMed PMID: 34349207; PubMed Central PMCID: PMCPMC8338973.

27. Gaertner K, Michell C, Tapanainen R, Goffart S, Saari S, Soininmaki M, et al. Molecular phenotyping uncovers differences in basic housekeeping functions among closely related

661 species of hares (Lepus spp., Lagomorpha: Leporidae). Mol Ecol. 2023;32(15):4097-117.
662 Epub 2022/11/03. doi: 10.1111/mec.16755. PubMed PMID: 36320183.
663 28. Meng G, Li Y, Yang C, Liu S. MitoZ: a toolkit for animal mitochondrial genome
664 assembly, annotation and visualization. Nucleic Acids Res. 2019;47(11):e63. Epub
665 2019/03/14. doi: 10.1093/nar/gkz173. PubMed PMID: 30864657; PubMed Central
666 PMCID: PMCPMC6582343.
667 29. Chen SF, Zhou YQ, Chen YR, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
668 Bioinformatics. 2018;34(17):884-90. doi: 10.1093/bioinformatics/bty560. PubMed
669 PMID: WOS:000444317200035.
670 30. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node
671 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
672 Bioinformatics. 2015;31(10):1674-6. Epub 2015/01/23. doi:
673 10.1093/bioinformatics/btv033. PubMed PMID: 25609793.
674 31. Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs.
675 Bioinformatics. 2013;29(19):2487-9. Epub 2013/07/12. doi:
676 10.1093/bioinformatics/btt403. PubMed PMID: 23842809; PubMed Central PMCID:
677 PMCPMC3777106.
678 32. Gertz EM, Yu YK, Agarwala R, Schaffer AA, Altschul SF. Composition-based statistics
679 and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC
680 Biol. 2006;4:41. Epub 2006/12/13. doi: 10.1186/1741-7007-4-41. PubMed PMID:
681 17156431; PubMed Central PMCID: PMCPMC1779365.
682 33. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. Genome Res.
683 2004;14(5):988-95. Epub 2004/05/05. doi: 10.1101/gr.1865504. PubMed PMID:
684 15123596; PubMed Central PMCID: PMCPMC479130.

685 34. Juhling F, Putz J, Bernt M, Donath A, Middendorf M, Florentz C, Stadler PF. Improved
686 systematic tRNA gene annotation allows new insights into the evolution of mitochondrial
687 tRNA structures and into the mechanisms of mitochondrial genome rearrangements.
688 Nucleic Acids Res. 2012;40(7):2833-45. Epub 2011/12/06. doi: 10.1093/nar/gkr1131.
689 PubMed PMID: 22139921; PubMed Central PMCID: PMCPMC3326299.
690 35. Bayona-Bafaluy MP, Manfredi G, Moraes CT. A chemical enucleation method for the
691 transfer of mitochondrial DNA to ρ° cells -: art. no. e98. Nucleic Acids Research.
692 2003;31(16). doi: ARTN e98
693 10.1093/nar/gng100. PubMed PMID: WOS:000184783000045.
694 36. Fontanesi L, Tazzoli M, Pecchioli E, Hauffe HC, Robinson TJ, Russo V. Sexing European
695 rabbits (Oryctolagus cuniculus), European brown hares (Lepus europaeus) and mountain
696 hares (Lepus timidus) with ZFX and ZFY loci. Mol Ecol Resour. 2008;8(6):1294-6. doi:
697 10.1111/j.1755-0998.2008.02167.x. PubMed PMID: WOS:000260528700022.
698 37. Melo-Ferreira J, Boursot P, Suchentrunk F, Ferrand N, Alves PC. Invasion from the cold
699 past: extensive introgression of mountain hare (Lepus timidus) mitochondrial DNA into
700 three other hare species in northern Iberia. Mol Ecol. 2005;14(8):2459-64. Epub
701 2005/06/23. doi: 10.1111/j.1365-294X.2005.02599.x. PubMed PMID: 15969727.
702 38. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-
703 quadruplexes in nucleotide sequences. Nucleic Acids Res. 2006;34(Web Server
704 issue):W676-82. Epub 2006/07/18. doi: 10.1093/nar/gkl253. PubMed PMID: 16845096;
705 PubMed Central PMCID: PMCPMC1538864.
706 39. Pham XH, Farge G, Shi Y, Gaspari M, Gustafsson CM, Falkenberg M. Conserved
707 sequence box II directs transcription termination and primer formation in mitochondria.
708 J Biol Chem. 2006;281(34):24647-52. Epub 2006/06/23. doi: 10.1074/jbc.M602429200.
709 PubMed PMID: 16790426.

40. Pohjoismaki JLO, Forslund JME, Goffart S, Torregrosa-Munumer R, Wanrooij S. Known Unknowns of Mammalian Mitochondrial DNA Maintenance. Bioessays. 2018;40(9):e1800102. Epub 2018/07/13. doi: 10.1002/bies.201800102. PubMed PMID: 29999547.

41. Wanrooij S, Fuste JM, Farge G, Shi Y, Gustafsson CM, Falkenberg M. Human mitochondrial RNA polymerase primes lagging-strand DNA synthesis in vitro. Proc Natl Acad Sci U S A. 2008;105(32):11122-7. Epub 2008/08/08. doi: 10.1073/pnas.0805399105. PubMed PMID: 18685103; PubMed Central PMCID: PMCPMC2516254.

42. Fuste JM, Wanrooij S, Jemt E, Granycome CE, Cluett TJ, Shi Y, et al. Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. Mol Cell. 2010;37(1):67-78. Epub 2010/02/05. doi: 10.1016/j.molcel.2009.12.021. PubMed PMID: 20129056.

43. Uliano-Silva M, Ferreira J, Krasheninnikova K, Darwin Tree of Life C, Formenti G, Abueg L, et al. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. BMC Bioinformatics. 2023;24(1):288. Epub 2023/07/19. doi: 10.1186/s12859-023-05385-y. PubMed PMID: 37464285; PubMed Central PMCID: PMCPMC10354987.

44. Melo-Ferreira J, Vilela J, Fonseca MM, da Fonseca RR, Boursot P, Alves PC. The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. Genome Biol Evol. 2014;6(4):886-96. Epub 2014/04/04. doi: 10.1093/gbe/evu059. PubMed PMID: 24696399; PubMed Central PMCID: PMCPMC4007550.

45. Stoddard BL. Homing endonucleases from mobile group I introns: discovery to genome engineering. Mob DNA. 2014;5(1):7. Epub 2014/03/05. doi: 10.1186/1759-8753-5-7. PubMed PMID: 24589358; PubMed Central PMCID: PMCPMC3943268.

46. Richard GF, Kerrest A, Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiol Mol Biol Rev. 2008;72(4):686-727. Epub 2008/12/05. doi: 10.1128/MMBR.00011-08. PubMed PMID: 19052325; PubMed Central PMCID: PMCPMC2593564.

47. Skipper KA, Andersen PR, Sharma N, Mikkelsen JG. DNA transposon-based gene vehicles - scenes from an evolutionary drive. J Biomed Sci. 2013;20(1):92. Epub 20131209. doi: 10.1186/1423-0127-20-92. PubMed PMID: 24320156; PubMed Central PMCID: PMCPMC3878927.

48. Yant SR, Kay MA. Nonhomologous-end-joining factors regulate DNA repair fidelity during Sleeping Beauty element transposition in mammalian cells. Mol Cell Biol. 2003;23(23):8505-18. doi: 10.1128/MCB.23.23.8505-8518.2003. PubMed PMID: 14612396; PubMed Central PMCID: PMCPMC262663.

49. Ogawa T, Yamada Y, Kuroda T, Kishi T, Moriya S. The datA locus predominantly contributes to the initiator titration mechanism in the control of replication initiation in Escherichia coli. Mol Microbiol. 2002;44(5):1367-75. Epub 2002/06/19. doi: 10.1046/j.1365-2958.2002.02969.x. PubMed PMID: 12068813.

50. Hansen FG, Christensen BB, Atlung T. The initiator titration model: computer simulation of chromosome and minichromosome control. Res Microbiol. 1991;142(2-3):161-7. Epub 1991/02/01. doi: 10.1016/0923-2508(91)90025-6. PubMed PMID: 1925015.

51. Hernández MA, Campos F, Gutiérrez-Corchero F, Amezcua A. Identification of Lanius species and subspecies using tandem repeats in the mitochondrial DNA control region. Ibis. 2004;146:227-30.

52. Herbers E, Kekalainen NJ, Hangas A, Pohjoismaki JL, Goffart S. Tissue specific differences in mitochondrial DNA maintenance and expression. Mitochondrion. 2019;44:85-92. Epub 2018/01/18. doi: 10.1016/j.mito.2018.01.004. PubMed PMID: 29339192.

53. Chretien D, Benit P, Ha HH, Keipert S, El-Khoury R, Chang YT, et al. Mitochondria are physiologically maintained at close to 50 degrees C. PLoS Biol. 2018;16(1):e2003992. Epub 2018/01/26. doi: 10.1371/journal.pbio.2003992. PubMed PMID: 29370167; PubMed Central PMCID: PMCPMC5784887.

54. Thulin C-G, Tegelström H. Biased geographical distribution of mitochondrial DNA that passed the species barrier from mountain hares to brown hares (genus Lepus): an effect of genetic incompatibility and mating behaviour? J Zool. 2002;258:299-306. doi: 10.1017/S0952836902001425. PubMed PMID: WOS:000179553600006.

Tables

Table 1. The mitochondrial genomes from mountain hare (LT) and brown hare (LE) cell lines sequenced in this study.

Cell line	Species	Country	Municipality	Size	GenBank
LT1	*Lepus timidus*	Finland	Ilomantsi	17,482 bp	OR915850
LT4	*Lepus timidus*	Finland	Vesilahti	17,483 bp	OR939641
LT5	*Lepus timidus*	Finland	Outokumpu	17,481 bp	OR939643
LT6	*Lepus timidus*	Finland	Ruokolahti	17,471 bp	OR939642
LE1	*Lepus europaeus*	Finland	Liperi	17,540 bp	OR915849
LE2	*Lepus europaeus*	Finland	Outokumpu	17,546 bp	OR876275
LE3	*Lepus europaeus*	Finland	Kontiolahti	17,541 bp	OR939639
LE4	*Lepus europaeus*	Finland	Vesilahti	17,578 bp	OR939640

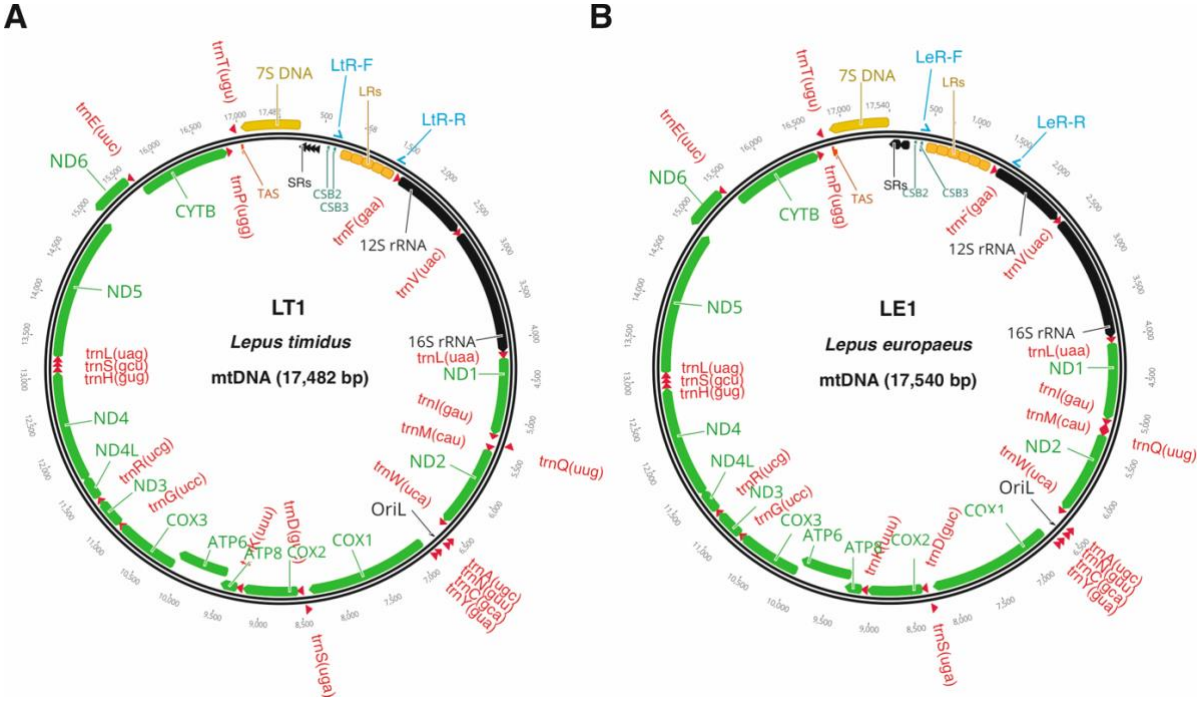
774 **Table 2.** The short repeat core element types and copy number variation among hares. Clade
775 indicates whether the species is more related to mountain hare (LT) or brown hare (LE). Core
776 element copy numbers represent only the analyzed mitochondrial genomes (see materials and
777 methods) and not a population study. LT+ for *Lepus yarkandensis* indicates a unique dimer
778 insertion in the classical LT-type SR (see text).

Species	**Species clade**	**SR type**	**Core element copy number**
Lepus timidus	LT	LT	32
Lepus tibetanus	LT	LT	16
Lepus tolai	LT	LT	17
Lepus arcticus	LT	LT	23
Lepus sinensis	LT	LT	28
Lepus coreanus	LT	LT	32
Lepus granatensis	LT	LT	32
Lepus otus	LT	LT	32
Lepus corsicanus	LT	LT	32
Lepus yarkandensis	LT	LT+	21
Lepus capensis	LE	LT	29
Lepus europaeus	LE	LE	28–32

779

780

781 **Figures**

782 

783 **Fig. 1.** Schematic illustration and gene maps of hare mitochondrial genomes. (A) The
784 mitochondrial DNA from LT1 mountain hare (*Lepus timidus*) and (B) LE1 brown hare (*L.*
785 *europaeus*) fibroblasts. The binding sites for the long tandem repeat (LR) genotyping primers
786 (LR-F, LR-R) are indicated. See Fig. 2 for the details of the non-coding region.

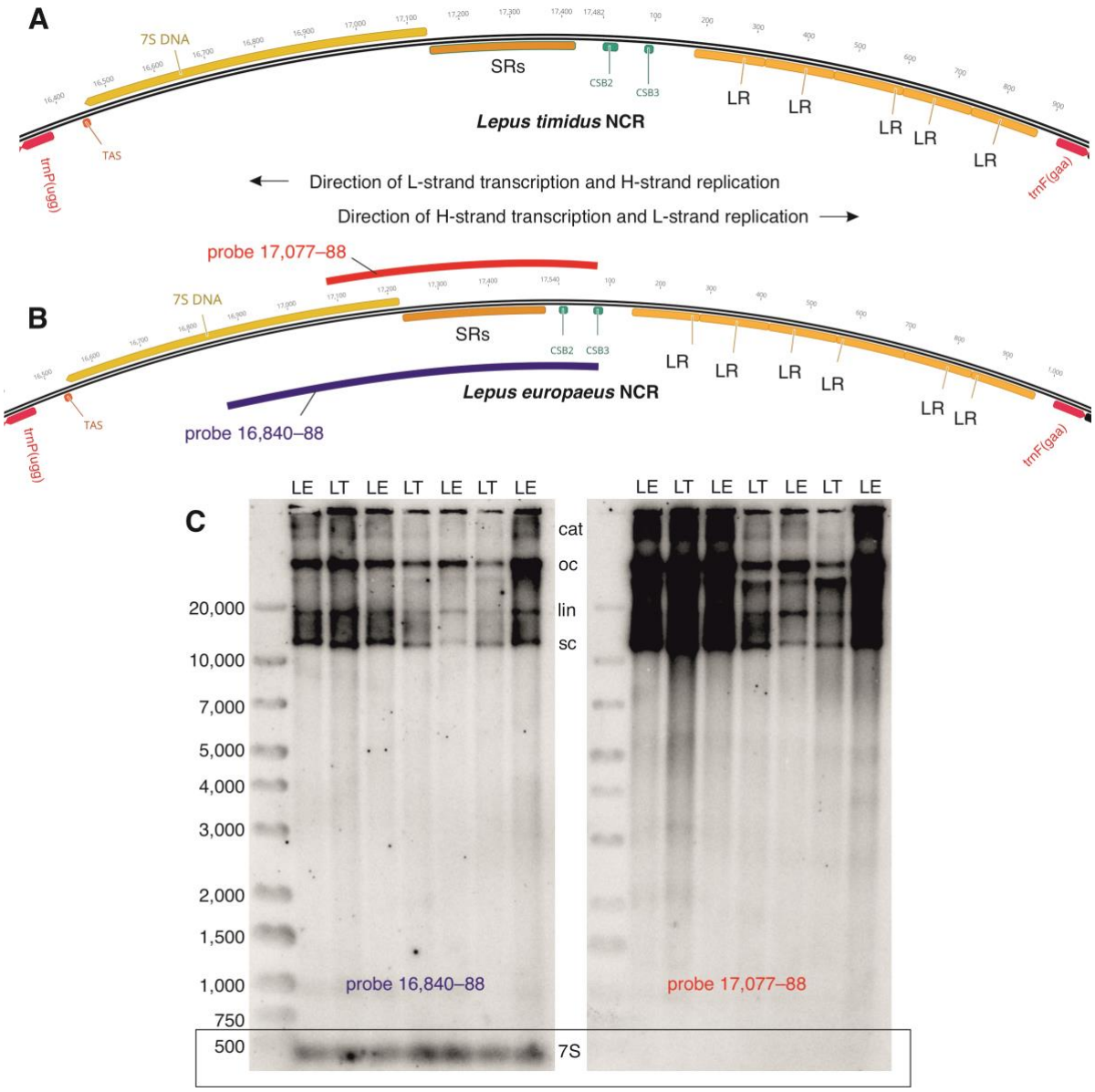


Fig. 2. Sequence features of hare mitochondrial DNA non-coding region (NCR). (A) The NCR of LT1 cells and (B) LE1 cells. The direction of transcription and replication, as well as probe locations for the Southern blot are indicated. Note that the H-strand replication is the leading-strand, initiated first and primed by the L-strand transcript. Key: trnP = tRNA-proline; TAS = termination associated sequence; SR = short tandem repeat; CSB2 and 3 = conserved sequence block 2 and 3; LR = long tandem repeat; trnaF = tRNA-phenylalanine. (C) A Southern blot of uncut mtDNA from mountain hare (LT) and brown hare cells (LE), probed for the region spanning nts 16,840–88 (left panel) or 17,077–88 (right panel). See the probe locations in (B).

Right panel is overexposed to demonstrate the absence of 7S signal. Key: cat = catenated mtDNA forms, oc = open circles; lin = linear mtDNA; sc = supercoiled circles; 7S = 7S DNA.

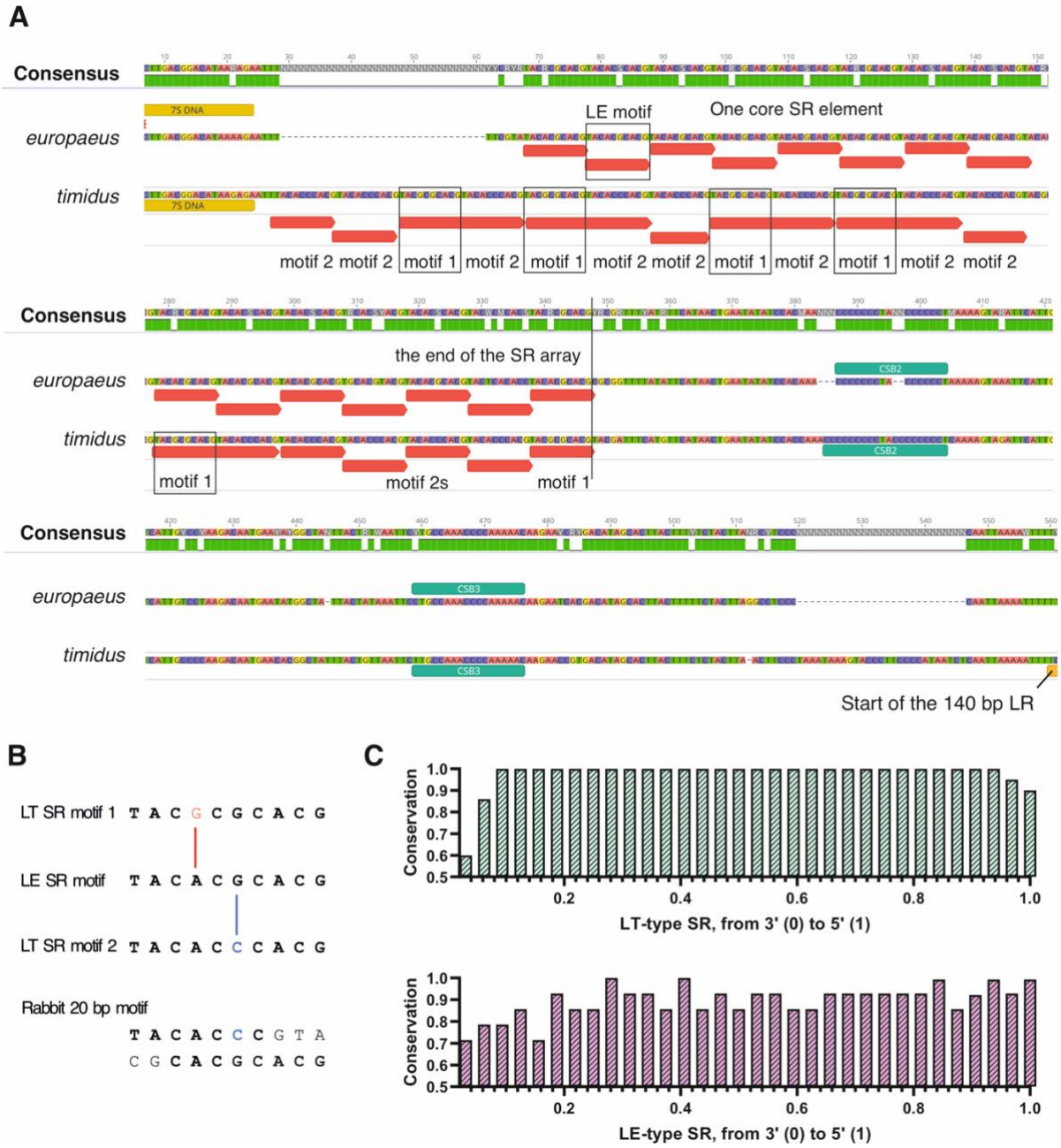


Fig. 3. Comparison of the sequence differences between the brown hare (*L. europaeus*) and mountain hare (*L. timidus*) mtDNA non-coding regions between the assumed start of the 7S DNA and the long tandem repeats. (A) The short tandem repeats (SRs) form a continuously repeating array of a single motif in brown hare, while the mountain hares show random

mixtures of two different repeat motifs. Repeating sequences and their direction are marked with red bars. Mountain hare CSB2 also has more Cs flanking the central TA-pair, whereas the CSB3 sequence is identical between the species. Note both SNP and indel variation between the two species. (B) The 10 bp conserved motif sequences in the hare SRs. LE-type is only seen in brown hares whereas LT-types are present in all hares analyzed here. Note the G-A transition (red) and G-C transversion (blue) between the three core sequences. The 20 bp rabbit SR motif is shown in two parts as comparison. Identical sequences in bold. Note the similarity of the two rabbit sequence halves with the brown hare repeat motif. (C) Sequence conservation of the core SR sequence motifs in hares. The LT-type shows remarkable conservation across species, whereas LE-type shows more variation within one species.

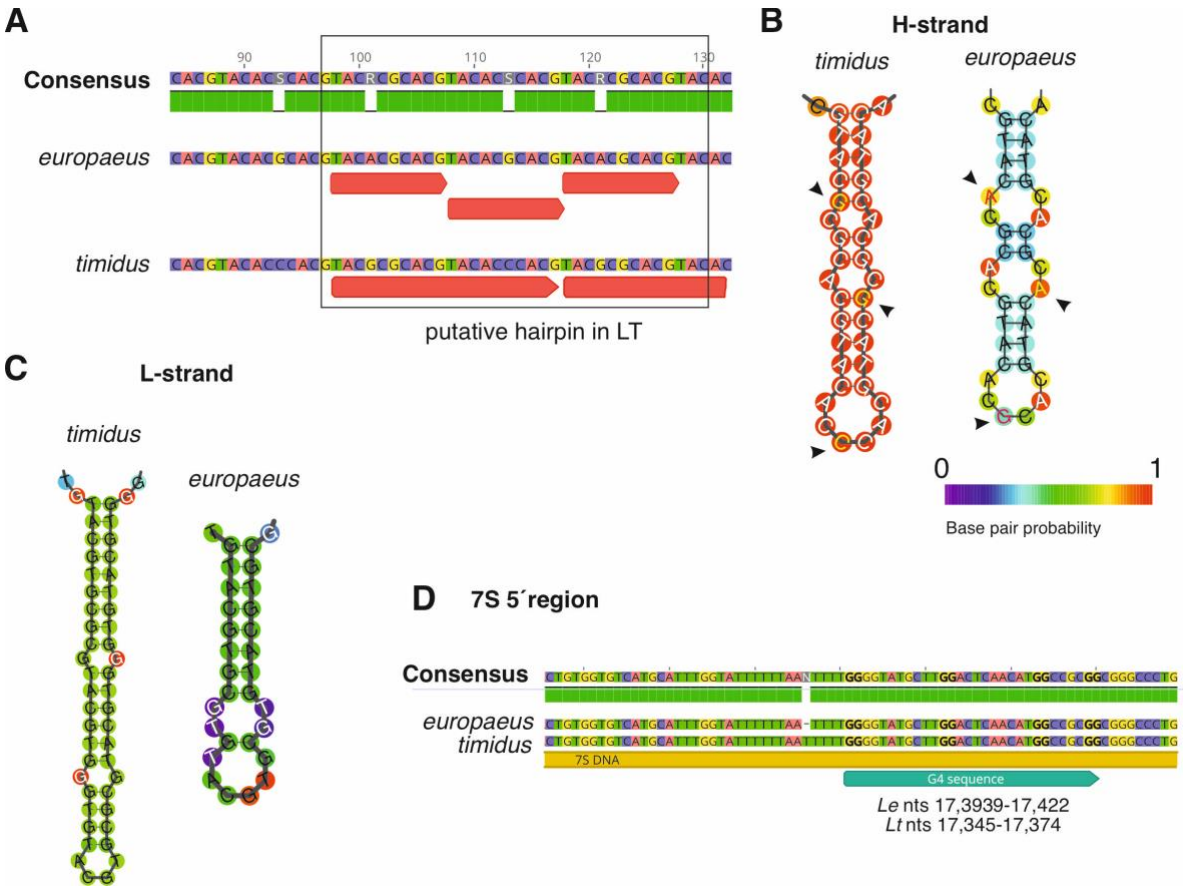
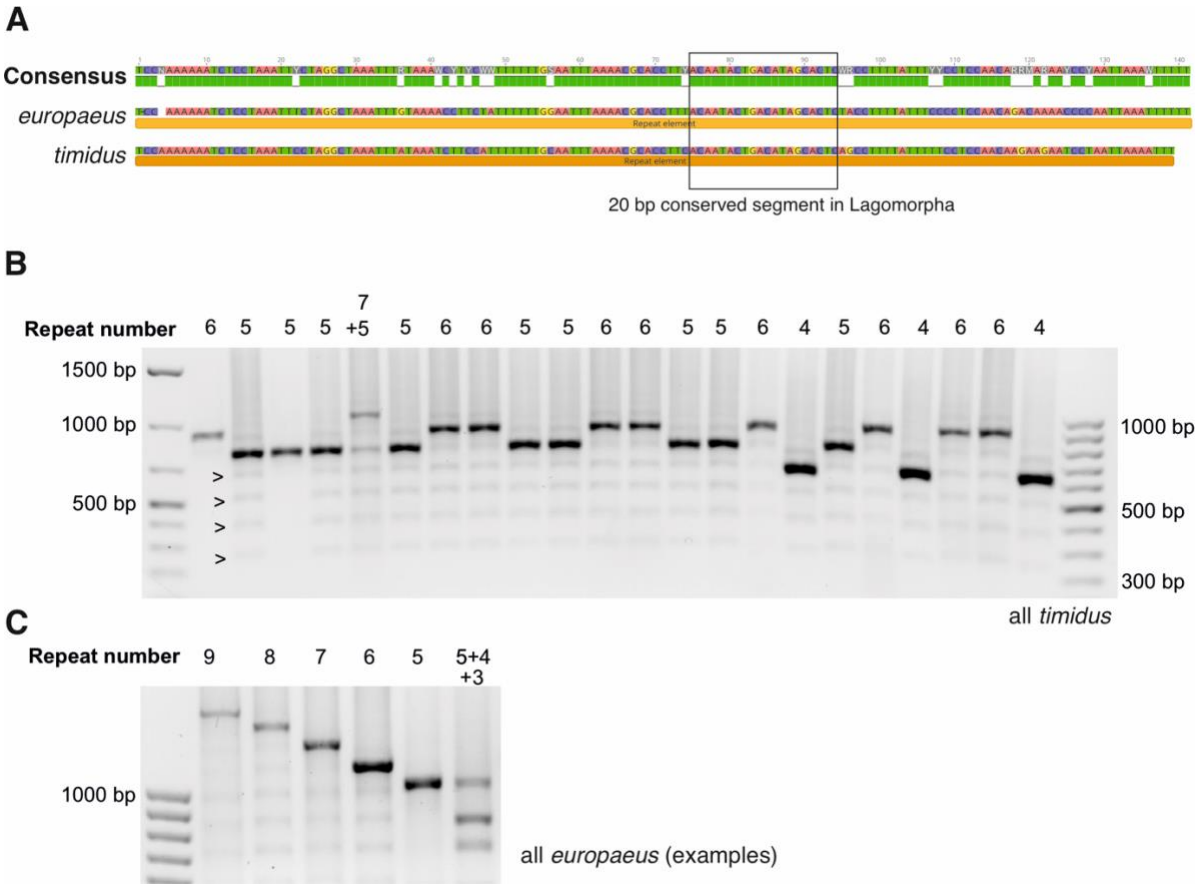


Fig. 4. Structural features of the hare mitochondrial non-coding region. (A) Despite the difference in the repeat element length and their overlap, there are only a few SNPs between the short repeat elements (SR) from the brown hare (*europaeus*) and mountain hare (*timidus*).

818 (B) Predicted mountain hare and brown hare SR H-strand secondary structures. Note the longer
819 and stronger hairpin formation in mountain hare. (C) Predicted secondary structures on the SR
820 L-strand. (D) A putative G-quadruplex sequence close to the 5´-end of the 7S DNA (nts 17,078–
821 17,107 in LE1). G-duplets are highlighted.



822

823 **Fig. 5.** Long tandem repeats (LR) in hares. (A) Sequence comparison of LRs from brown hares
824 and mountain hares, highlighting the 20 bp conserved sequence present in all known
825 Lagomorpha. (B) An example of LR copy number variation in mountain hares. Arrowheads
826 highlight LR ladder generated as a PCR artefact. (C) Examples of LR copy number variants
827 present in brown hares. Size difference compared to (B) due to different PCR primers.

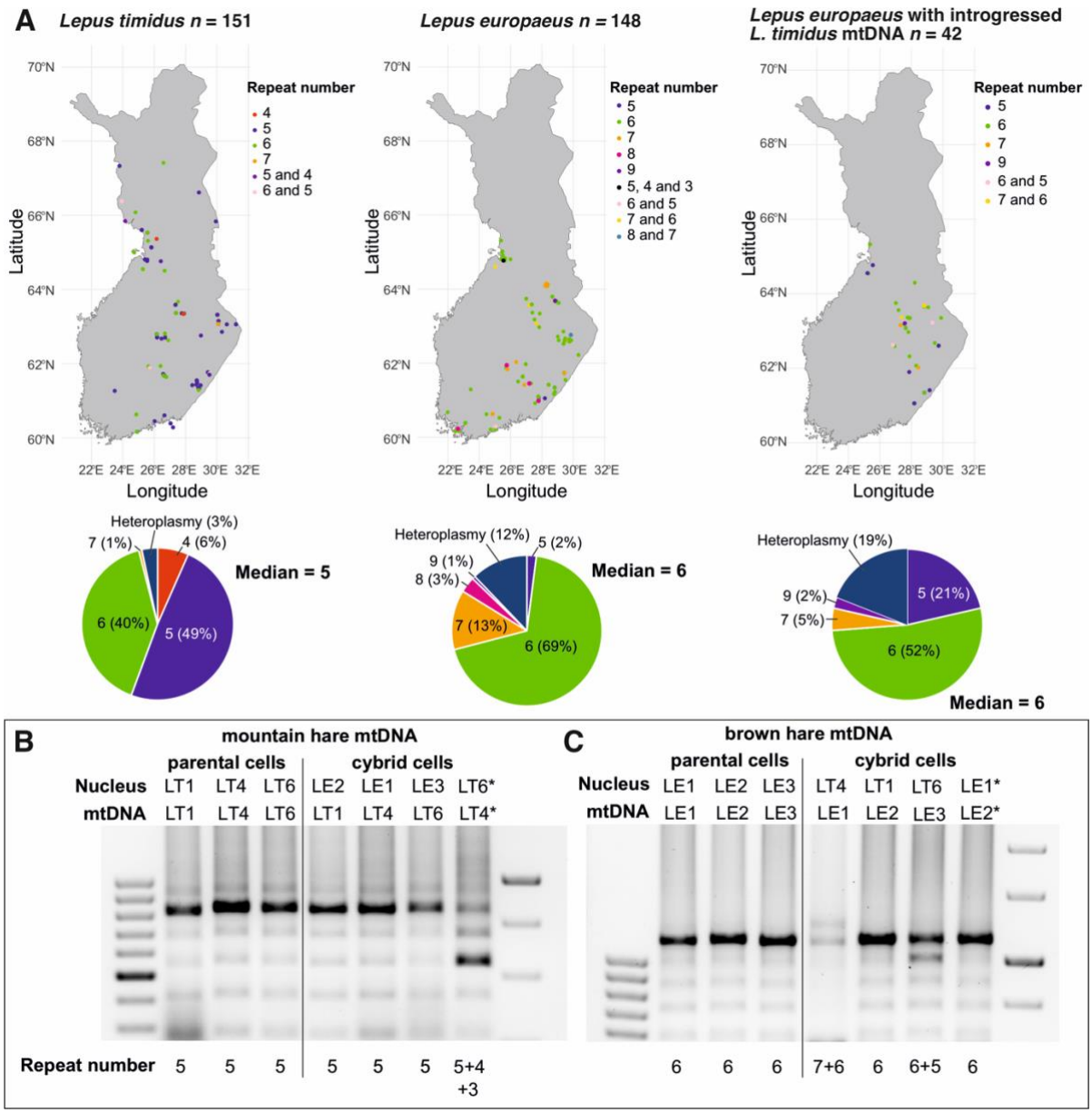


Fig. 6. Long tandem repeat (LR) copy number variation in Finnish hares. (A) The geographical distribution of LR copy number variants and their frequencies in Finnish mountain hares (*Lepus timidus*) and brown hares (*L. europaeus*) with conspecific or introgressed mtDNA. (B) LR copy numbers in the original mountain hare fibroblasts and xenomitochondrial cybrids with mtDNA from these cells. (C) The same for the parental and cybrid cells with brown hare mtDNA. LT = *Lepus timidus*; LE = *L. europaeus*. Asterisk indicates control cybrids with nucleus and mtDNA from the same species, but different parental cell lines.