

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

4

5      **Riikka Tapanainen<sup>1</sup>, Koit Aasumets<sup>1</sup>, Zsófia Fekete<sup>1,2</sup>, Steffi Goffart<sup>1</sup>, Eric Dufour<sup>3¶</sup>, &**  
6      **Jaakko L. O. Pohjoismäki<sup>1\*¶</sup>**

7

8

9      <sup>1</sup>University of Eastern Finland, Department of Environmental and Biological Sciences –  
10     Joensuu, Finland

11     <sup>2</sup>Hungarian University of Agriculture and Life Sciences, Institute of Genetics and  
12     Biotechnology, Gödöllő, Hungary

13     <sup>3</sup>Mitochondrial Bioenergetics and Metabolism, Faculty of Medicine and Health Technology,  
14     FI-33014 Tampere University, Finland

15

16     \*Corresponding author: [Jaakko.Pohjoismaki@uef.fi](mailto:Jaakko.Pohjoismaki@uef.fi)

17     ¶These authors contributed equally

18

19

20

21

22

## 23 Abstract

24 The non-coding regions of the mitochondrial DNAs (mtDNAs) of hares, rabbits, and pika  
25 (Lagomorpha) contain short (~20 bp) and long (130–160 bp) tandem repeats, absent in related  
26 mammalian orders. In the presented study, we provide functional annotation for mountain hare  
27 (*Lepus timidus*) and brown hare (*L. europaeus*) mtDNA non-coding regions, together with a  
28 species- and population-level analysis of tandem repeat variation. Mountain hare short tandem  
29 repeats (SRs) as well as other analyzed hare species consist of two conserved 10 bp motifs,  
30 with only brown hares exhibiting a single, more variable motif. Long tandem repeats (LRs)  
31 also differ in sequence and copy number between species. Mountain hares have four to seven  
32 LRs, median value five, while brown hares exhibit five to nine LRs, median value six.  
33 Interestingly, introgressed mountain hare mtDNA in brown hares obtained an intermediate LR  
34 length distribution, with median copy number being the same as with conspecific brown hare  
35 mtDNA. In contrast, transfer of brown hare mtDNA into cultured mtDNA-less mountain hare  
36 cells maintained the original LR number, whereas the reciprocal transfer caused copy number  
37 instability, suggesting that cellular environment rather than the nuclear genomic background  
38 plays a role in the LR maintenance. Due to their dynamic nature and separation from other  
39 known conserved sequence elements on the non-coding region of hare mitochondrial genomes,  
40 the tandem repeat elements are unlikely to have regulatory roles but are likely to represent  
41 selfish genetic elements.

42

43 **Keywords:** non-coding region; regulatory element; mitochondrial DNA replication; strand-  
44 slippage; population genetics; mountain hare; brown hare, repeat element

45

## 46 1 Introduction

47 Mitochondria are essential cellular organelles best known for the production of ATP through a  
48 process called oxidative phosphorylation (OXPHOS) (Spinelli and Haigis, 2018). They are also  
49 a central hub for most metabolic pathways and exhibit specialized functions such as the  
50 biosynthesis of iron-sulfur clusters, steroids, fatty acid oxidation, calcium homeostasis,  
51 programmed cell death, and more. Due to their origin as a free-living prokaryote, mitochondria  
52 have their own small circular DNA genome (mitochondrial DNA, mtDNA), which has  
53 undergone reductive evolution associated with gene transfer to the nucleus (Adams and Palmer,  
54 2003). In animals, mtDNA is typically less than 20 kb in size and contains genes for 13  
55 polypeptides of the OXPHOS complexes as well as two mitochondrial ribosomal RNAs and  
56 22 tRNAs required for their translation (Boore, 1999). The two strands of the mtDNA have  
57 very different gene density, with most of the genes being encoded on the so-called heavy strand  
58 (H-strand). The designation stems from its high guanine content, giving it a higher mass  
59 compared to the complementary light-strand (L-strand), as visualized in alkaline CsCl density  
60 gradient centrifugation (Kasamatsu et al., 1971). Besides the highly compact coding region,  
61 the mitochondrial genome also has a non-coding region (NCR), sometimes called the control  
62 region due to the presence of control elements required for DNA transcription and replication.  
63 The length and organization of the NCR is very variable across the animal kingdom,  
64 contributing most to the mtDNA length variation between the species (Boore, 1999).

65 In vertebrates, the mtDNA non-coding region contains interesting features, such as the  
66 displacement loop (D-loop), which is generated by replication initiating from the main  
67 replication origin (OriH: Origin of Heavy strand replication) and terminating prematurely at  
68 the “termination associated sequence” (TAS) (Jemt et al., 2015). As a result, a triple-stranded  
69 DNA structure is formed, including a 600–700 nt long nascent H-strand fragment, the 7S DNA,  
70 which remains hybridized to its L-strand template. While the human and mouse mitochondrial

71 genomes, probably the most studied mtDNAs, show considerable economy in their NCR  
72 organization, its size can be variable in other vertebrates due to tandem repeat arrays (Lunt et  
73 al., 1998). Of these, short tandem repeats of around 10 bp are an universal feature for the insect  
74 mitochondrial genomes (Solignac et al., 1986), but are also present in vertebrate species,  
75 including mammals (Savolainen et al., 2000). Longer, 100-200 bp repeats are found in certain  
76 species of fish (Wang et al., 2007), birds (Mundy and Helbig, 2004; Omote et al., 2013; Wang  
77 et al., 2015) and some mammals, such as rabbits and hares (Casane et al., 1997). The repeat  
78 number can vary from two (Wang et al., 2007) to more than one hundred (Hoelzel, 1993) per  
79 mtDNA molecule. As tandem repeats can be easily lost during the sequencing and assembly of  
80 short sequence reads, such as the ones obtained from Illumina sequencing platforms, it is likely  
81 that they are an overlooked feature of mitochondrial genome variation across different taxa.  
82 For example, an analysis involving long read sequencing data from vertebrate genome  
83 assemblies discovered tandem repeats and gene duplications from several species (Formenti et  
84 al., 2021). Interestingly, the presence of tandem repeats was not systematic, but they seem to  
85 have been lost and obtained independently in different evolutionary lineages.

86 The tandem repeats on the NCR of the European rabbit (*Oryctolagus cuniculus*)  
87 mitochondrial genome were noted already in the 1980s (Ennafaa et al., 1987) and later work  
88 showed that these repeats are widespread also in other mitochondrial genomes of Lagomorpha  
89 (Casane et al., 1997). There are two types of repeated motifs in rabbits, 20 bp short repeats  
90 (SR) and 153 bp long repeats (LR). SR arrays are expanded or contracted relatively  
91 dynamically, with the copy number of the units varying between three to 19. Consequently, the  
92 SR arrays can occur in heteroplasmy, i.e. mitochondrial DNAs with different repeat numbers  
93 coexisting in the same cell or tissue. An elegant experiment where rabbit SRs were cloned and  
94 maintained in a bacterial plasmid demonstrated that slipped-strand mispairing during  
95 replication is the main mechanism to explain the dynamic state of the repeat array (Pfeuty et

96 al., 2001). The experiments demonstrated that SR insertions are more common than deletions  
97 in arrays with less than 10 copies, with the opposite being true for longer tracts, resulting in the  
98 SR lengths to oscillate around an optimal value.

99 Similar to the SRs, the LRs also have a high mutation rate ( $10^{-2}$  per animal per  
100 generation), resulting in NCR length variation between generations as well as within  
101 individuals, manifesting as heteroplasmy as well as mosaicism (Casane et al., 1997). In rabbits,  
102 LR tend to be present as arrays of five repeats with shorter and longer variants being rarer.  
103 While the mtDNA haplotype has no influence of the array length, oddly, some organs such as  
104 gonads maintain longer arrays than other tissues (Casane et al., 1997). Due to their length, the  
105 expansion or deletion of LRs through replication slippage-mispairing is more complicated than  
106 for the SRs. In the nucleus, the copy number changes of similar long satellite repeats, including  
107 expansions and contractions of rDNA arrays (Kobayashi and Ganley, 2005), have been  
108 proposed to occur through break-induced recombination (Thakur et al., 2021). Interestingly, a  
109 similar copy-choice recombination mechanism involving strand invasion of the parental  
110 molecule by a free 3'-end of a newly synthesized DNA strand at regions of short sequence  
111 homologies has been suggested as a mechanism for mtDNA deletion formation in humans  
112 (Persson et al., 2019; Phillips et al., 2017). Non-allelic recombination or end joining could in  
113 principle also explain the observed LR length variation.

114 LR tandem repeats have been found in all Lagomorphs from pikas (*Ochotona* spp.) to  
115 hares (*Lepus* spp., *Sylvilagus* spp.) (Casane et al., 1997; Dufresne et al., 1996). All LRs have a  
116 20 bp strictly conserved sequence, which has been proposed to be a binding site for a regulatory  
117 factor, such as TFAM (Dufresne et al., 1996). In fact, the rabbit LR has been suggested to  
118 contain both L-strand (LSP) and H-strand (HSP) promoters for mtDNA transcription (Dufresne  
119 et al., 1996). If this is the case, variation in LR length could influence primer synthesis by

120 MTRPOL at these promoters (Kuhl et al., 2016), thereby modulating also the replication  
121 initiation.

122 Our group has been interested in hybridization between the mountain hare (*Lepus*  
123 *timidus*) and brown hare (*L. europaeus*) at the northern contact zone of these species in Finland.  
124 Despite being separated by three million years of evolution (Ferreira et al., 2021), the hybrids  
125 are fertile, resulting in gene flow across the species barrier (Levanen et al., 2018a; Levanen et  
126 al., 2018b). This gene flow is biased towards the brown hare, which seems to benefit from the  
127 hybridization by obtaining locally adapted alleles from the mountain hare (Pohjoismaki et al.,  
128 2021). Furthermore, the hybridization results in frequent introgression of mountain hare  
129 mtDNA into the brown hare population, encompassing up to fifth of the individuals in certain  
130 regions in Finland (Levanen et al., 2018a). When sequencing the mitochondrial genomes of  
131 our mountain hare and brown hare cell lines (Gaertner et al., 2023), we recently noted a species  
132 difference in the length of LRs in the mtDNA non-coding region. A population sampling of 151  
133 mountain hares and 148 brown hares with species-specific mtDNA confirmed that brown hares  
134 maintain more LR copies (median = 6) than mountain hares (median = 5), with no correlation  
135 with their geographic origin. Interestingly, when introgressed into brown hares, the mountain  
136 hare mtDNA presented frequent heteroplasmy and gained “brown-hare like” LRs (median = 6  
137 copies). To further test how the nuclear background can influence LRs, we generated cybrids  
138 of mountain hare and brown hare fibroblasts with nuclear DNA from one species and mtDNA  
139 from another. The repeat region length changed in three cybrid cell lines out of six, regardless  
140 of their genetic background, with the new LR variant maintained in heteroplasmy. In one case,  
141 the LR number of brown hare mtDNA was increased from six to seven in cells with mountain  
142 hare nucleus, a state which is very rare in the natural population of mountain hares. We  
143 conclude that while the nuclear genetic background certainly plays a role in the LR  
144 maintenance, environmental factors and differences in cell biology between species also

145 contribute to the observed variation. We also discuss the potential relation of the NCR repeat  
146 elements to the regulation of mtDNA.

147 **2 Materials and Methods**

148 **2.1 Sampling and DNA isolation**

149 The mountain hare (*Lepus timidus*) and brown hare (*L. europaeus*) specimens presented here  
150 are from a larger biobank of 1,202 hare samples collected in 2012–2016 across Finland  
151 (Levanen et al., 2018a; Levanen et al., 2018b). The origin and generation of the four mountain  
152 hare (LT1, LT4, LT5, LT6) and four brown hare (LE1, LE2, LE3, LE4) fibroblast cell lines has  
153 been described elsewhere (Gaertner et al., 2023). Most samples are from hunted animals, with  
154 a subset from specimens found dead from the wild. The sampling had minimal impact on the  
155 local populations and posed no threat on the habitats. The hunting followed the regional hunting  
156 seasons and legislation (Metsästyslaki [Hunting law] 1993/615/5§). The sampling adhered to  
157 the ARRIVE guidelines and no ethical assessment was required. All sampling occurred in  
158 Finland and did not involve International Trade in Endangered Species of Wild Fauna and Flora  
159 (CITES) or other export of specimens, as defined by the Convention on Biological Diversity  
160 (CBD). The DNA was isolated from frozen ear muscle biopsies using the peqGOLD blood &  
161 tissue DNA mini kit (VWR) and following the protocol provided by the manufacturer.

162

163 **2.2 Mitochondrial DNA sequencing**

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

165 Le93F: TTGTTTTGTAGCAAGTTACACATGC

166 Le184R: GCTTAATACCTGCTCCTCTTGATCTA

167 Le1580F: TTAAACCCATAGTTGGCCTAAAAGC

168 Le1635R: TTGAGCTTAAACGCTTCTTAATTGA

169 Le3045F: AGGCGTATTATTTATCCTAGCAACCT  
170 Le3175R: CCTCATAAGAAATGGTCTGTGCGA  
171 Le3921F: CCCCCTAATCTTCCATCATCCTAT  
172 Le4482R: TCATCCTATATGGCAATTGAGGAAT  
173 Le4689F: AGGCTTATTCCAAAGTGAATTATTATTCA  
174 Le5417R: AGGCTCCAAATAAAAGGTAGAGAGTT  
175 Le6696F: ATACCGTCTCATCAATAGGCTCCTC  
176 Le6756R: ATAAAGATTATTACTATTACAGCGGTTAGA  
177 Le8603F: AGCCTATATCTACATGATAACTTAATGA  
178 Le8698R: CGGATAAGGCCCGGTAAGTGG  
179 Le10552F: TTGAAGCAACACTAATCCCTACACTA  
180 Le10613R: TCGTTCTGTTGATTACCTCATCGT  
181 Le11301F: ACCATTAACCTCTAGGAGAGCTTCT  
182 Le11807R: AGGATAATGATTGAGACGGCTATTGA  
183 Le12407F: GTCTAACCTAGCTGCTACAGGTAAG  
184 Le12791R: GAGCATAAAAAGAGTATAGCTTGAA  
185 Le14204F: ATTGTTAACCACTCTCTAACGACCT  
186 Le14514R: CCAATGTTCAGGTTCTAGGTAAGT  
187 Lt16056F: TGGGGTATGCTTGGACTCAAC  
188 Le16119R: TCGTCTACAATAAGTGCACCGG  
189  
190 In total, 12 separate reactions were prepared to cover the mitochondria genome:  
191 1. Lt16056F + Le184R: 1871 bp  
192 2. Le93F + Le1635R: 1543 bp  
193 3. Le1580F + Le3175R: 1596 bp

194 4. Le3045F + Le4482R: 1438 bp

195 5. Le3921F + Le5417R: 1497 bp

196 6. Le4689F + Le6756R: 2068 bp

197 7. Le6696F + Le8698R: 2003 bp

198 8. Le8603F + Le10613R: 2011 bp

199 9. Le10552F + Le11807R: 1256 bp

200 10. Le11301F + Le12791R: 1491 bp

201 11. Le12407F + Le14514R: 2108 bp

202 12. Le14204F + Le16119R: 1916 bp

203 (Expected fragment size based on the published *Lepus europaeus* mtDNA sequence from  
204 Sweden [NC\_004028.1]).

205

206 The fragments were amplified from total DNA preparations using a PCR program with a 1 min  
207 94 °C denaturing step, followed by 35 cycles of 94 °C for 15 s, 56 °C for 15 s and 72 °C for 2  
208 min and a final 3 min elongation step at 72 °C. The obtained products were gel purified using  
209 the GeneJET gel extraction kit (Thermo Fisher Scientific™) and sent for sequencing using  
210 Illumina MiniSeq™ at the Genome Center of Eastern Finland.

211

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

215 Le101F: TATAAATTCTGCCAACCCCAAAAA

216

217 **2.3 Mitochondrial DNA assembly and annotation**

218 Mitochondrial DNA was assembled using the MitoZ pipeline(Meng et al., 2019) and the best  
219 assembly was selected by comparing the results of the pipeline's outputs with different kmer  
220 options. The final assembly was done using the un options --clade Chordata –fastq\_read\_length  
221 150, --requiring\_taxa Chordata --genetic\_code 2 --kmers\_megahit 21 29 39 59 79 99 119 141.  
222 The pipeline included the tools fastp (Chen et al., 2018) for cleaning the raw data, MEGAHIT  
223 (Li et al., 2015) for assembly, after which sequences were filtered using HMMER (Wheeler  
224 and Eddy, 2013) to ensure the correct taxa and the completeness of protein-coding genes.  
225 Annotation was performed using TBLASTN (Gertz et al., 2006), GeneWise (Birney et al.,  
226 2004) and MiTFi (Juhling et al., 2012). The annotation of the non-coding region (NCR) as well  
227 as the illustration of the mitochondrial genome was done with Geneious® 10.2.6 (Biomatters.  
228 Available from <https://www.geneious.com>). The functional loci on the NCR were identified  
229 based on the similarity with the human (NC\_012920) and mouse (FJ374652) NCR sequences.  
230

231 **2.4 Genotyping of the long repeat (LR) region**

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

234 *Lepus timidus* mtDNA

235 LtLR-F: AGAACCGTGACATAGCACTTACTTTC

236 LtLR-R: TAACATATTGGTGTAGAATGTTTTAGT

237 *L. europaeus* mtDNA

238 LeLR-F: TATAAATTCCTGCCAAACCCCAAAAA

239 LeLR-R: GCTTAATACCTGCTCCTCTTGATCTA

240

241 The used PCR program had an initial denaturation step of 94 °C for 2 min, followed by 35  
242 cycles of 94 °C for 20 s, 59 °C for 20 s and 72 °C for 90 s, with final elongation at 72 °C for  
243 5 min, using AccuStart II™ PCR SuperMix (Quantabio). The PCR products were separated  
244 over a 1 % TAE agarose gel.

245

## 246 **2.5 Analyses of the SRs**

247 Core SR sequences and their flanking regions of the Finnish *Lepus timidus* and *Lepus*  
248 *europaeus* were manually retrieved from the whole mitochondrial genome sequences. All  
249 possible core dimers (1 for LE type, 4 for LT) were then used as query sequence for a basic  
250 nucleotide BLAST search. Results were filtered to only include mitochondrial genomes  
251 sequences from defined geographic isolates (LE query) or from one genome per *Lepus* species  
252 (LT queries). For LT queries; *Lepus tibetanus*: MN539746.1, *L. tolai*: MN539744.1, *L. arcticus*:  
253 NC\_044769.1, *L. sinensis*: NC\_025316.1, *L. coreanus*: NC\_024259.1, *L. granatensis*:  
254 NC\_024042.1, *L. othus*: KJ397608.1, *L. corsicanus*: KJ397606.1, *L. capensis*: NC\_015841.1  
255 and *L. yarkandensis*: MN539747.1 mitochondrial genomes were selected. *L. townsendii* was  
256 excluded due to poor quality sequencing of its SR region. For the LE query, *L. europaeus*  
257 Poland 3: KY211034.1, Poland 2: KY211033.1, Poland 1: KY211032.1, Germany 2:  
258 KY211031.1, Germany 1: KY211030.1, Greece 4: KY211029.1, Greece 3: KY211028.1,  
259 Greece 1: KY211026.1, Cyprus 4: KY211025.1, Cyprus 3: KY211024.1, Cyprus 2:  
260 KY211023.1, Cyprus 1: KY211022.1, and Turkey 1: KY211021.1 were selected. LE Greece 2  
261 isolate was excluded as its SR region was identical to that of Greece 3 isolate. Rabbit  
262 (*Oryctolagus cuniculus*) reference genome mtDNA (NC\_001913.1) was used for comparison.  
263 The proportion of mutated core repeats was compared using Fisher exact test. Because of its  
264 unique SR features, *L. yarkandensis* was excluded from the statistical analyses. Sequence  
265 uncertainties in the repeats of *L. otus* (1) and *L. corsicanus* (3) were all considered as mutated

266 for the Fisher exact test, but the mutation frequency was nevertheless significantly higher in  
267 the LE type of SRs. To visualize the frequency of mutation in function of the repeat location,  
268 the repeat positions were normalized to the number of repeats in the SR leading to an x-axis  
269 ranging from 0 (“beginning” of the SR region) to 1 (“end” of the SR region).

270

## 271 **2.6 Generation of cybrid cell lines**

272 Cybrid cells were generated using the chemical enucleation method (Bayona-Bafaluy  
273 et al., 2003). In brief, first recipient mtDNA-less ρ0 cells from immortalized LT1, LT4, LT6,  
274 LE1, LE2 and LE3 fibroblast cell lines were generated by culturing them in the presence of  
275 100 µM ddC, until no mtDNA was detected on a Southern blot. The ρ0 cells were maintained  
276 in high glucose (4.5 g/l) DMEM, supplemented with 10 % FBS and 50 µg/ml uridine. The  
277 nucleus of the mtDNA donor cells was removed with Actinomycin D, with the effective  
278 concentration determined separately for each cell line. The optimal concentration for most cell  
279 lines was 1 µg/ml for 18 h, except for LT1 (1 µg/ml for 15 h) and LE2 (2 µg/ml for 28 h). The  
280 cybrid fusions were conducted in both directions (same cell line either a recipient or a donor  
281 of mtDNA) for the following cell line pairs: LE1♂ x LT4♀, LE2♀ x LT1♂ and LE3♀ x LT6♂.  
282 For the control cybrids with mtDNA from the same species but a different cell line than the  
283 nucleus, we used LE1♂(nucleus) x LE2♀(mtDNA) and LT4♀ (mtDNA) x LT6♂ (nucleus)  
284 combinations. The fusion was performed by growing the mtDNA donor cells on 6-well plates  
285 and treating them with Actinomycin D. After replacing the treatment medium with fresh  
286 DMEM containing 10 % FBS 1 million ρ0 cells were added to the well for 3 h. The cells were  
287 then washed 3 × with fresh DMEM before addition of 45 % polyethylene glycol (MW 1450  
288 g/mol) for 60 s. The cells were then washed again 3 × with DMEM + 10 % DMSO and once  
289 with DMEM. Finally, the cells were given DMEM + 10% FBS + Penicillin/Streptomycin.  
290 Unlike the ρ0, the cybrid cells were maintained without added uridine to select for functional

291 mitochondrial DNA. The cells were grown to confluence, at which stage half of the cells in a  
292 well were collected for genotyping. DNA was extracted with Quick-DNA Miniprep Kit (Zymo  
293 Research).

294 As the donor and recipient cells differed by their sex, PCR-RFLP of *ZFX* and *ZFY* loci  
295 (Fontanesi et al., 2008) was used to genotype the cybrid nucleus and *CYTB* PCR-RFLP their  
296 mtDNA (Melo-Ferreira et al., 2005). Each original cybrid pool was genotyped and promising  
297 looking pools were subcloned on 96-wells. These clonal cell lines were rechecked for the  
298 expected nuclear and mtDNA haplotypes, obtaining the final cybrid cell lines.

299

### 300 **2.7 Agarose gel electrophoresis and Southern hybridization**

301 Mitochondrial DNA from LE and LT parental and cybrid cell lines was isolated using  
302 phenol:chloroform extraction and ethanol precipitation. 2 µg of uncut mitochondrial DNA were  
303 separated over a 0.4 % agarose gel in 1x TBE buffer at 25 V overnight at room temperature. To  
304 facilitate the transfer of mtDNA onto nylon membrane by capillary transfer, the gel was  
305 subjected to acid depurination followed by denaturation to create suitable hybridization targets.  
306 The transferred DNA was crosslinked by baking 80 °C for 2 h. The Southern blot was probed  
307 using ( $\alpha$ -32P)-dCTP labelled PCR probe spanning the hare mtDNA nucleotides 16840-88 or  
308 17077-88 at 65 °C for overnight. Signals were captured on a phosphor screen and detected  
309 using a phosphorimager (Fujifilm FLA-3000).

310

### 311 **2.8 Statistical tests**

312 The RNAfold web server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) for  
313 predicting secondary structures of single stranded RNA or DNA sequences was used to screen  
314 for hairpin formation in the mitochondrial non-coding region and the QGRS mapper (Kikin et  
315 al., 2006) to identify putative Quadruplex forming G-Rich Sequences.

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

319 **3 Results**

320 **3.1 The structure of the mitochondrial DNA non-coding region in hares**

321 We amplified the mitochondrial genomes from four mountain hare (LT1, LT4, LT5, LT6) and  
322 four brown hare (LE1, LE2, LE3, LE4) cell lines as 12 overlapping PCR products, which were  
323 subsequently sequenced using Illumina Miniseq™. When assembling the sequences, we noted  
324 that the obtained sequence for the NCR was shorter than expected from the PCR product  
325 lengths. Closer inspection revealed that this was because the short sequence reads ( $2 \times 150$  bp)  
326 over the LR region were assembled as a single repeat element due to their redundancy. The  
327 PCR products containing NCR elements were therefore reanalyzed using Sanger sequencing  
328 to clarify LR copy number and finalize the assemblies (Table 1, Fig. 1). Interestingly, our  
329 mountain hare as well as brown hare mitochondrial genomes are smaller (Table 1) than the  
330 previously published mountain hare mtDNA from a Finnish specimen (NC\_024040.1) (17,755  
331 bp) or the brown hare mtDNA from Sweden (NC\_004028.1). The size difference is explained  
332 by the length of the non-coding region (NCR), where our LT cell lines have five and LE cell  
333 lines six long tandem repeats (LRs) compared to seven in both previously published  
334 mitochondrial genomes.

335

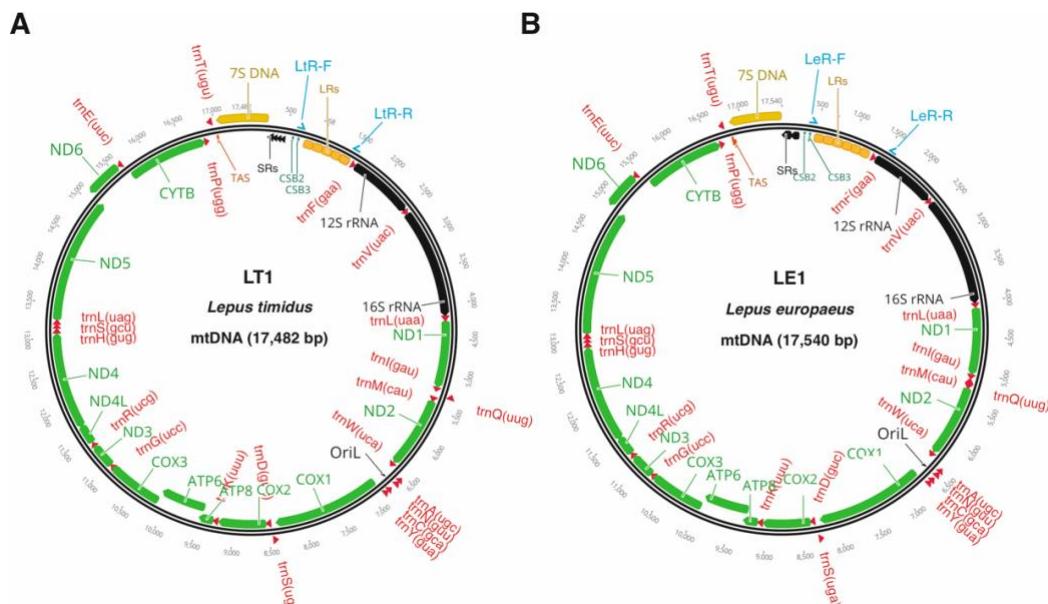
336 **Table 1.** The mitochondrial genomes from mountain hare (LT) and brown hare (LE) cell lines  
337 sequenced in this study.

| Cell line | Species              | Country | Municipality | Size      | GenBank  |
|-----------|----------------------|---------|--------------|-----------|----------|
| LT1       | <i>Lepus timidus</i> | Finland | Ilomantsi    | 17,482 bp | OR915850 |

|     |                        |         |             |           |          |
|-----|------------------------|---------|-------------|-----------|----------|
| LT4 | <i>Lepus timidus</i>   | Finland | Vesilahti   | 17,483 bp | OR939641 |
| LT5 | <i>Lepus timidus</i>   | Finland | Outokumpu   | 17,481 bp | OR939643 |
| LT6 | <i>Lepus timidus</i>   | Finland | Ruokolahti  | 17,471 bp | OR939642 |
| LE1 | <i>Lepus europaeus</i> | Finland | Liperi      | 17,540 bp | OR915849 |
| LE2 | <i>Lepus europaeus</i> | Finland | Outokumpu   | 17,546 bp | OR876275 |
| LE3 | <i>Lepus europaeus</i> | Finland | Kontiolahti | 17,541 bp | OR939639 |
| LE4 | <i>Lepus europaeus</i> | Finland | Vesilahti   | 17,578 bp | OR939640 |

338

339

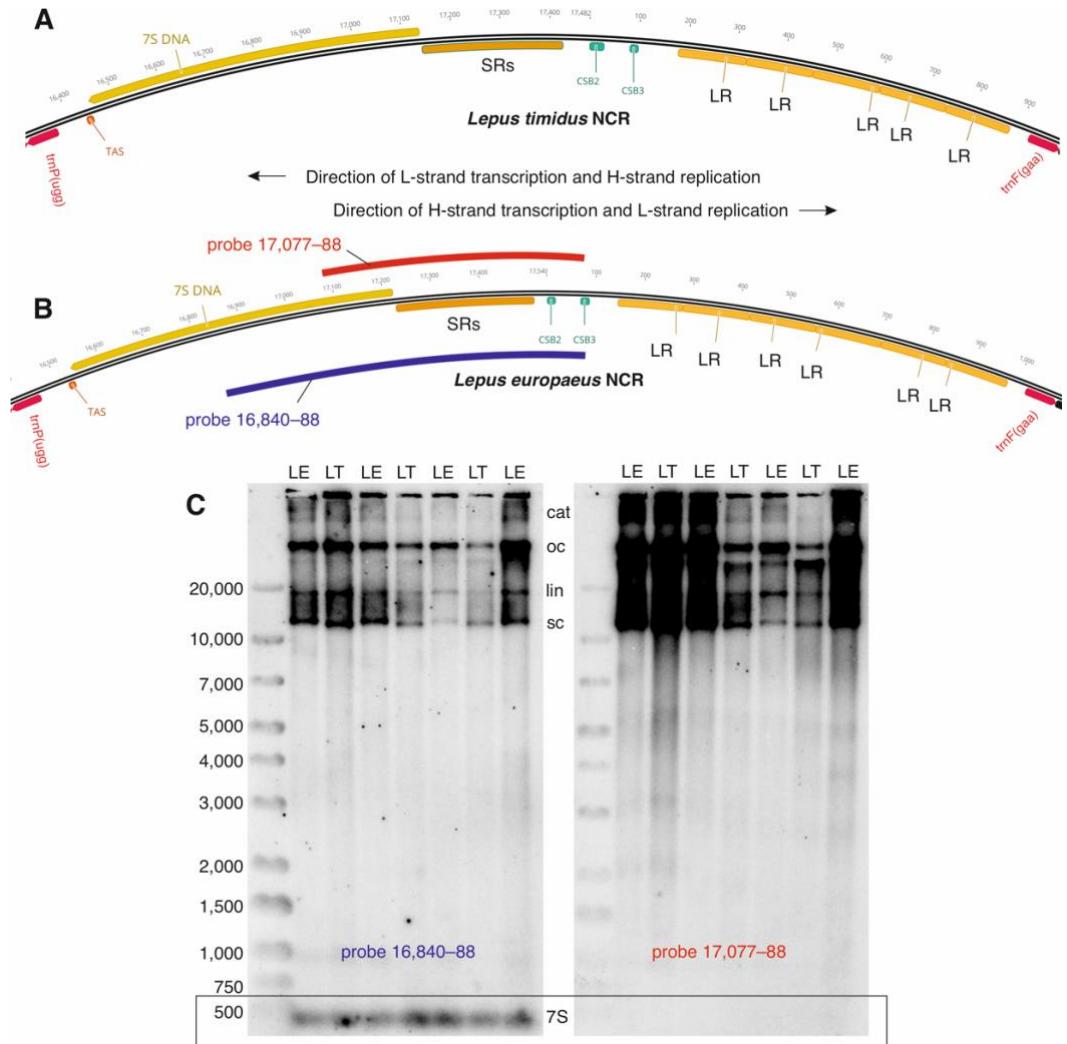


340

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

346

347 The length of the NCR is 2,038 bp in LT1 (Fig. 2A) and 2,102 bp in LE1 (Fig. 2B). It  
348 contains recognizable features such as a rather conserved region between the termination  
349 associated sequence (TAS) and the cluster of short tandem repeats (SR), which by length and  
350 location could correspond the D-loop region containing the 7S DNA. The SRs and the LRs are  
351 separated by a variably long sequence containing the conserved sequence blocks (CSBs) 2 and  
352 3 thought to be relevant for the control of mtDNA gene expression and replication priming  
353 (Pham et al., 2006). However, if replication is primed at or near the CSBs, the primer 3'-ends  
354 are relatively far away from the known 5'-ends of DNA, considered to correspond the  
355 replication origin and start of the 7S DNA (Pohjoismaki et al., 2018). In fact, Southern blot  
356 analysis of this region can only detect the 7S DNA when using a probe that extends downstream  
357 of the SRs (Fig. 2C). Also, the size of the 7S does not differ between the cell lines despite the  
358 difference in the length of the NCR-region. Of note, we could confirm that Lagomorphs do not  
359 seem to have any CSB1(Casane et al., 1997; Dufresne et al., 1996). The CSB2 differs by a  
360 couple of indels between the species, while the CSB3 seems highly conserved (Fig. 3A).



**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). The right panel is overexposed to demonstrate the absence

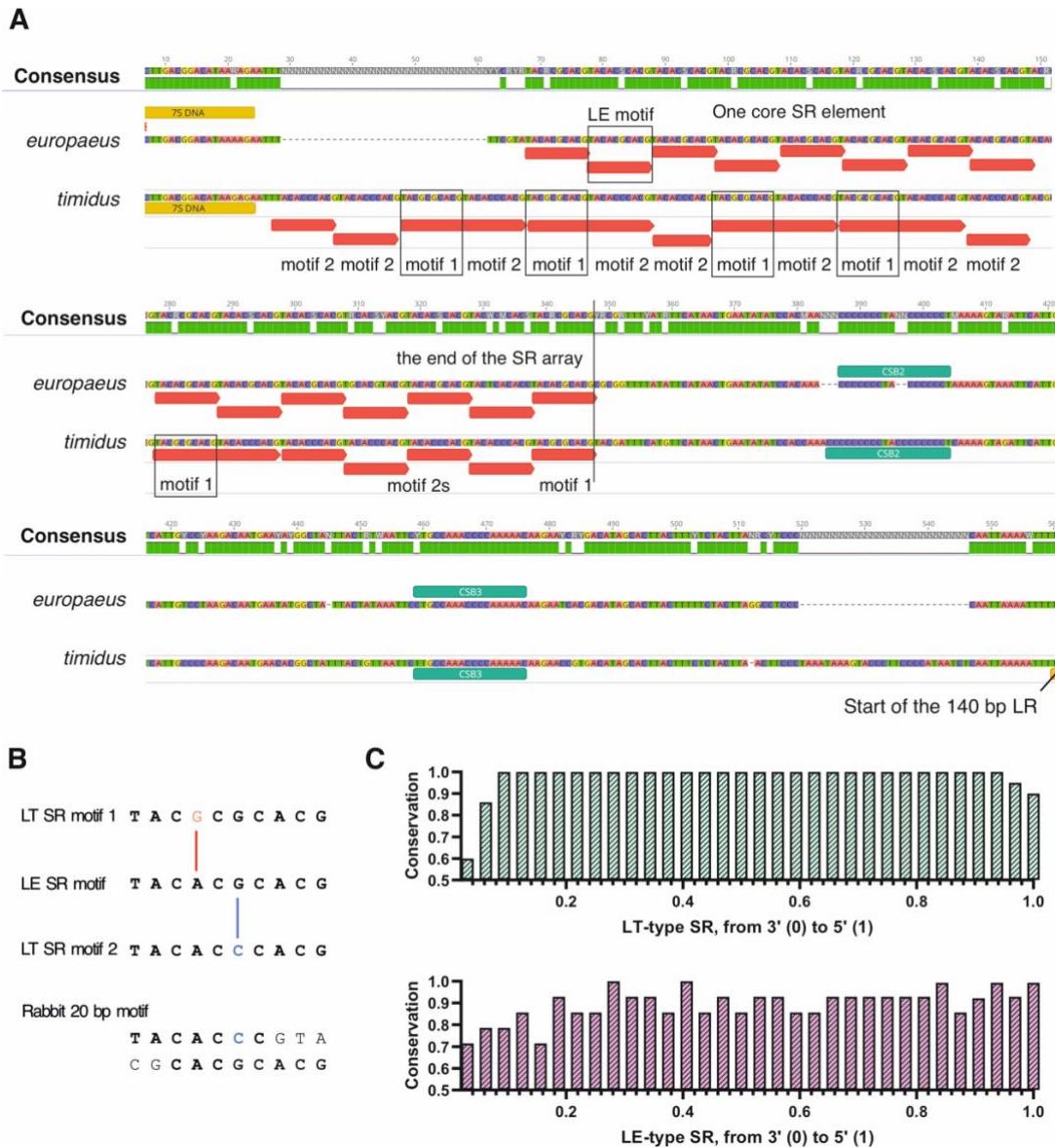
372 of 7S signal. Key: cat = catenated mtDNA forms, oc = open circles; lin = linear  
373 mtDNA; sc = supercoiled circles; 7S = 7S DNA.

374

### 375 **3.2 Hare SR elements consist of 10 bp core motifs**

376 At first glance, the mountain hare (LT) SRs appear as a 20 bp repeat element similar in length  
377 to the rabbit's SRs (Fig. 3A). However, these are not consistent, and part of the SR arrays  
378 showed repetition of only half of the sequence. On closer inspection, the element consists of  
379 two 10 bp motifs, TACCGCGCACG, and TACACCCCACG, which differ by the two underlined  
380 nucleotides. In contrast, the brown hare (LE) SRs contain only one type of 10 bp core motif  
381 (TACACGCCACG), which differs from both LT motifs by one of the variable nucleotides (Fig.  
382 3B). Interestingly, the 20 bp rabbit SR motif contains the 7 first nucleotides of the brown hare  
383 (LE) SR motif spliced together with the last 8 nucleotides of LT SR motif1, separated by an  
384 unique 5 bp (GTACG) spacer sequence (Fig. 3B).

385 While the LE type is present only in brown hares, the two LT types occur in all other  
386 hare species that we analyzed (Table 2), with the two core elements alternating variably in the  
387 SR arrays. Curiously, also the cape hare (*Lepus capensis*), a close relative of brown hare  
388 (Ferreira et al., 2021), has LT type SRs. *Lepus yarkandensis* constitute another curiosity, with  
389 randomly alternating core 1 and 2 LT type SR motifs being separated by dinucleotides (TT or  
390 TC). Overall, the LT type SRs show considerably less variation than the LE types (Fig. 3C,  
391 3.35%;  $p < 0.001$ , Fisher's exact test), despite being spread across many species. The copy  
392 number of the LT core repeat sequences varied from 32 in our LT cell lines and other *L. timidus*,  
393 *coreanus*, *granatensis*, *otus* and *corsicanus* isolates to 16 in *L. tibetanus* (Table 2). The copy  
394 number of the LE type, present only in brown hares, ranged from 28 to 32.



395

396 **Fig. 3.** Comparison of the sequence differences between the brown hare (*L.*  
 397 *europaeus*) and mountain hare (*L. timidus*) mtDNA non-coding regions between the  
 398 assumed start of the 7S DNA and the long tandem repeats. (A) The short tandem  
 399 repeats (SRs) form a continuously repeating array of a single motif in brown hare,  
 400 while the mountain hares show random mixtures of two different repeat motifs.  
 401 Repeating sequences and their direction are marked with red bars. Mountain hare  
 402 CSB2 also has more Cs flanking the central TA-pair, whereas the CSB3 sequence  
 403 is identical between the species. Note both SNP and indel variation between the two  
 404 species. (B) The 10 bp conserved motif sequences in the hare SRs. The LE type is

405 only seen in brown hares whereas LT types are present in all hares analyzed here.

406 Note the G-A transition (red) and G-C transversion (blue) between the three core  
407 sequences. The 20 bp rabbit SR motif is shown in two parts as comparison. Identical  
408 sequences in bold. Note the similarity of the two rabbit sequence halves with the  
409 brown hare repeat motif. (C) Sequence conservation of the core SR sequence motifs  
410 in hares. The LT type shows remarkable conservation across species, whereas the  
411 LE type shows more variation within one species.

412

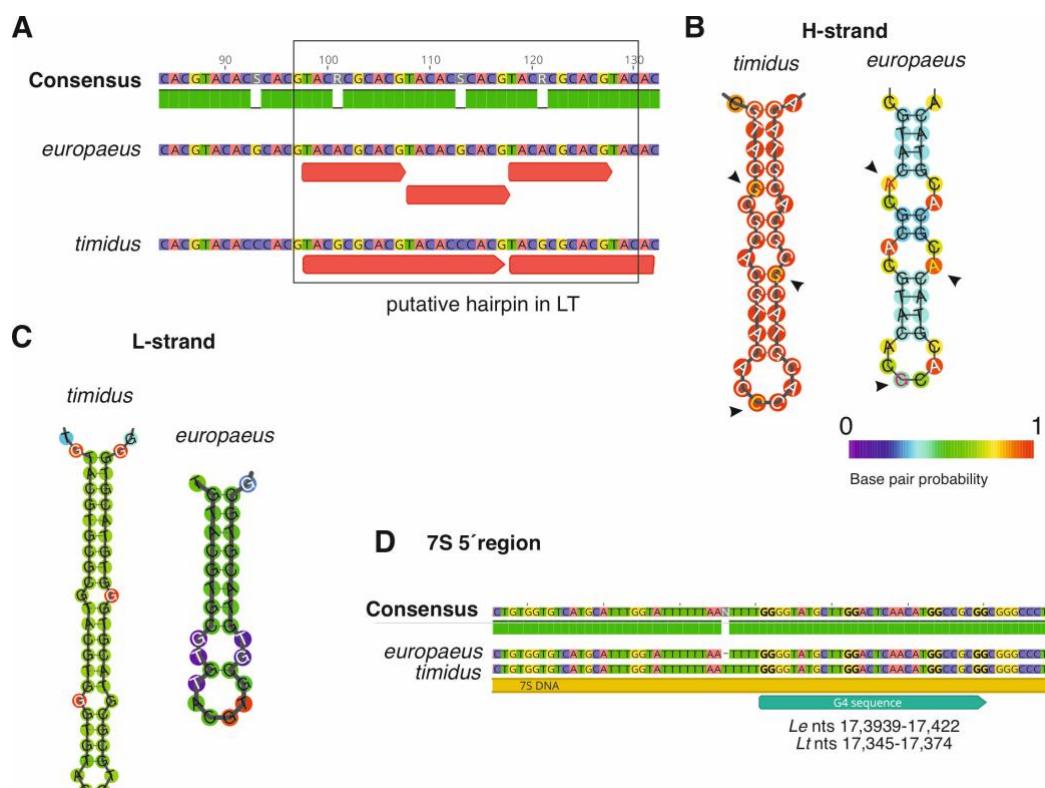
413 **Table 2.** The short repeat core element types and copy number variation among hares. The  
414 clade indicates whether the species is closer related to mountain hare (LT) or brown hare (LE).  
415 Core element copy numbers represent only the analyzed mitochondrial genomes (see materials  
416 and methods) and not a population study. LT+ for *Lepus yarkandensis* indicates a unique dimer  
417 insertion in the classical LT type SR (see text).

| Species                   | Species clade | SR type | Core element copy number |
|---------------------------|---------------|---------|--------------------------|
| <i>Lepus timidus</i>      | LT            | LT      | 32                       |
| <i>Lepus tibetanus</i>    | LT            | LT      | 16                       |
| <i>Lepus tolai</i>        | LT            | LT      | 17                       |
| <i>Lepus arcticus</i>     | LT            | LT      | 23                       |
| <i>Lepus sinensis</i>     | LT            | LT      | 28                       |
| <i>Lepus coreanus</i>     | LT            | LT      | 32                       |
| <i>Lepus granatensis</i>  | LT            | LT      | 32                       |
| <i>Lepus otus</i>         | LT            | LT      | 32                       |
| <i>Lepus corsicanus</i>   | LT            | LT      | 32                       |
| <i>Lepus yarkandensis</i> | LT            | LT+     | 21                       |
| <i>Lepus capensis</i>     | LE            | LT      | 29                       |

|                        |    |    |       |
|------------------------|----|----|-------|
| <i>Lepus europaeus</i> | LE | LE | 28–32 |
|------------------------|----|----|-------|

418

419 As repetitive sequences are likely to form secondary structures, we were interested to  
420 see whether the SR sequences of the two hares can form hairpins when single-stranded. This is  
421 particularly interesting as hairpin formation at the second most prominent replication origin on  
422 mitochondria, the light-strand origin (OriL) located at the WANCY tRNA cluster, is proposed  
423 to be required for L-strand replication priming (Fuste et al., 2010; Wanrooij et al., 2008). A  
424 rather strong hairpin was predicted for the H-strand SRs from the mountain hares but not from  
425 the brown hares, nor on the L-strand SR sequences of either species (Fig. 4A–C). As for other  
426 structural features, the entire non-coding region contained only a single sequence stretch at the  
427 5' end of the 7S DNA predicted to form a strong guanine quadruplex (G4) (Fig. 4D).



428

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

432 mountain hare (*timidus*). (B) Predicted mountain hare and brown hare SR H-strand  
433 secondary structures. Note the longer and stronger hairpin formation in mountain  
434 hare. (C) Predicted secondary structures on the SR L-strand. (D) A putative G-  
435 quadruplex sequence close to the 5' end of the 7S DNA (nts 17,078–17,107 in LE1).  
436 G-duplets are highlighted.

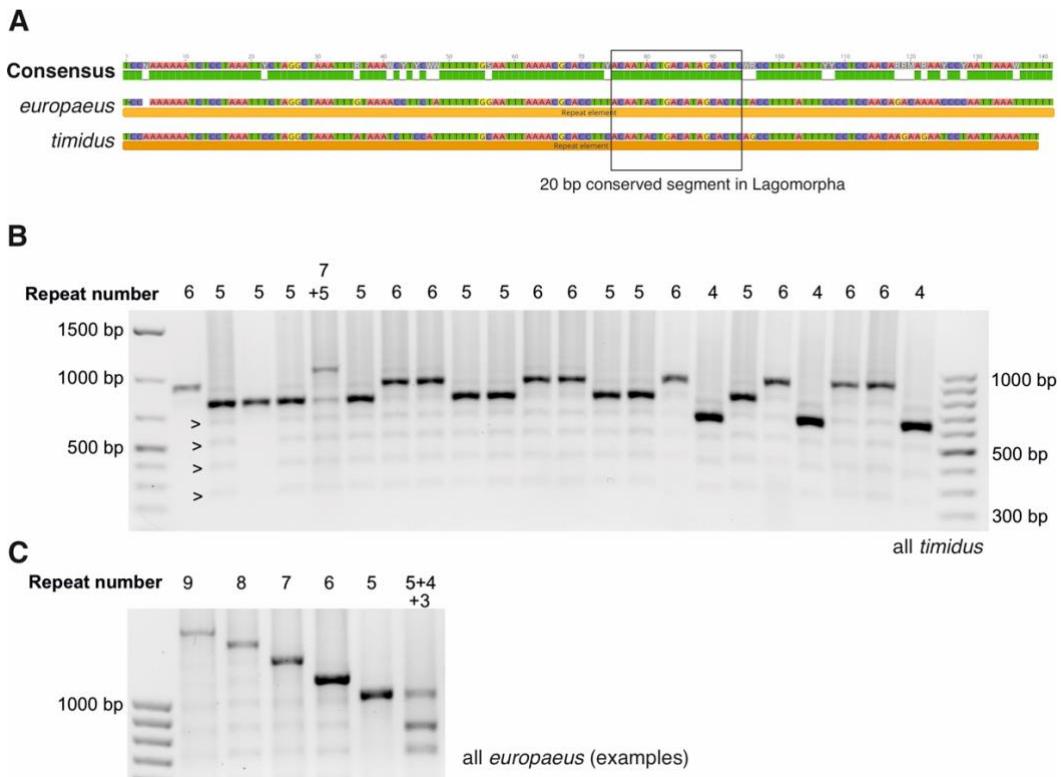
437

438 **3.3 LR length variation in Finnish hares**

439 The LRs of the two species are similar in length (Fig. 5A), 139 bp in the mountain hare and  
440 140 bp in the brown hare, differing at 21 nucleotide locations and containing the same 20 bp  
441 conserved sequence element as previously reported for all Lagomorpha (Casane et al., 1997).  
442 As also noted earlier, the LR copy number shows variation between and even within individuals  
443 (Fig. 5B, C). Although PCR tends to cause the amplification of short repeat sequences as an  
444 artefact, which are evident as a faint ladder of PCR products of varying repeat content, the  
445 common repeat haplotype was readily detectable as a strong band in all population samples  
446 and systematically confirmed the sequencing results.

447 While the mechanism of dynamic length variation in the SRs has been shown to result  
448 from replication slippage (Pfeuty et al., 2001), the factors influencing the LR array length are  
449 less clear. As they are also known to have species differences (Lunt et al., 1998), we were  
450 interested to see whether the LR array length would show population-specific variation. The  
451 most common length of LR was five repeats in Finnish mountain hares, with six repeats being  
452 the next abundant (Fig. 6A). The shortest observed repeat was four and longest seven, which  
453 was represented only by one sample. Curiously, the only previously published mountain hare  
454 mtDNA from Finland (NC\_024040.1) also has seven repeats. Heteroplasmy was detected in  
455 five samples (3 %). No geographic pattern in the LR length variation was observed among the

456 151 mountain hares across Finland (Moran I statistic standard deviate =  $8.09 \times 10^{-10}$ , p-value =  
457 0.5).



458

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

465

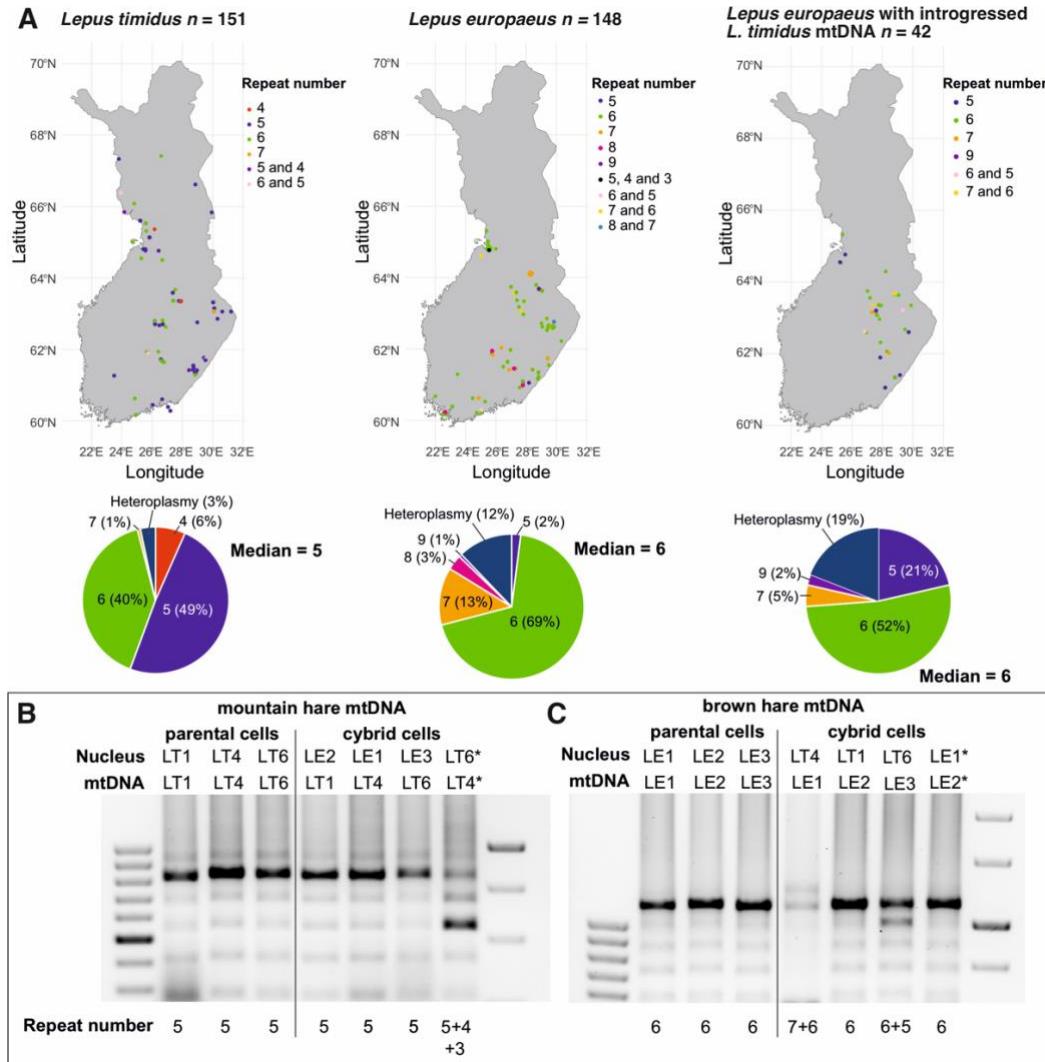
466 In contrast to mountain hares, the most common LR region length in brown hares was  
467 six repeats, with seven and eight repeats being more common than five (Fig. 6A). The longest  
468 observed LR region had nine repeats, while shorter than five repeats occurred only in  
469 heteroplasmic samples (Fig. 5C). In the heteroplasmic samples, the shortest repeat sequence  
470 length was three copies. Overall, brown hares showed more frequent heteroplasmy (12 %,  $p =$

471 0.0045, Fisher exact test). Interestingly, brown hares with introgressed mountain hare mtDNA  
472 showed an intermediate LR-region length distribution mtDNA (Fig. 6A). Although they had  
473 the same median length of six repeats as brown hares with conspecific mtDNA, five repeats  
474 were much more common (21 % vs. 2 %). Also, a notable proportion (19 %) of the introgressed  
475 samples were heteroplasmic, although the difference to the other brown hares was not  
476 significant ( $p = 0.3077$ , Fisher exact test). Again, no correlation between the LR length and  
477 geography was observed in brown hares with or without introgressed mtDNA. Only a few  
478 mountain hares with brown hare mtDNA were present in our collection (Levanen et al., 2018a)  
479 and therefore were not included in the analysis.

480

#### 481 **3.4 Transfer of mtDNA between cells causes instability in the LR copy number**

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



493

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

502

503 **4 Discussion**

504 When sequencing the mitochondrial DNAs as a part of the basic characterization of our hare  
505 fibroblast cell lines (Gaertner et al., 2023), we noted a discrepancy between the expected  
506 genome size and the assemblies obtained from short-read Illumina data. Closer inspection  
507 revealed that this was because of misassembly of the part of the non-coding region with known  
508 long tandem repeats (LRs). We corrected the assemblies with Sanger sequencing and provide  
509 here the fully annotated mitochondrial genomes for four mountain hares and four brown hares  
510 from Finland (Table 1, Figs. 1–2). Although tandem repeats are a known elements of the non-  
511 coding region in many vertebrate species, including lagomorphs, they remain an overlooked  
512 aspects of mitochondrial genome variation, largely due to their absence in humans and other  
513 well-studied model organisms. In addition, and as in our case initially, they are likely to be lost  
514 when assembling mitochondrial genome sequencing data obtained using short-read sequencing  
515 platforms. For this reason, a lot of hare mitochondrial genome assemblies available in GenBank  
516 have only one LR-element and many of the SR-array sequences might have issues as well. It  
517 is likely that the same problem applies to the mitochondrial genome assemblies from other  
518 organisms, meaning that the frequency and diversity of such elements in vertebrates is probably  
519 underestimated. Recent development of long-molecule sequencing technologies can help to  
520 complete our understanding of the repeat element variation, especially as the full-length  
521 mitochondrial genomes can be assembled reliably from whole-genome sequencing data  
522 (Uliano-Silva et al., 2023).

523 Not much is known about the function and biological significance of the repeat  
524 elements in the non-coding region of mitochondrial genomes. Initial studies proposed that the  
525 Lagomorph LRs are involved in the regulation of transcription and replication initiation  
526 (Casane et al., 1997; Dufresne et al., 1996). However, there are other conserved regulatory  
527 elements on the NCR (Fig. 2), which are known to function in e.g. mtDNA replication priming

528 in other mammals (Pham et al., 2006). Furthermore, the 7S DNA, resulting from prematurely  
529 terminated replication, does not contain sequences between LR- and SR-regions (Fig. 2C),  
530 indicating that the replication is likely to start at the SR, as suggested by other authors (Melo-  
531 Ferreira et al., 2014).

532 Interestingly, the SR array length can change rather dynamically in rabbits (Casane et  
533 al., 1997; Dufresne et al., 1996; Pfeuty et al., 2001) and their copy number is highly variable  
534 (from three to 19). In hares, the nature of the repeating sequences made the delimitation of the  
535 actual repeat element difficult, as different types of repeats with varying length can be identified  
536 (Fig. 3A). Like in rabbits, a 20 bp SR element can be identified in mountain hares. However,  
537 this element consists in fact of two separate motifs (Fig. 3B), which can occur also alone as the  
538 repeating element in the SR array. In brown hares, the SR arrays consist only of one core 10 bp  
539 motif, which differs from both mountain hare motif types by a single nucleotide (Fig. 3B). If  
540 only these canonical sequences are counted, their copy number in the different hare species  
541 varied between 16 and 32 (Table 2). Considering the evolution of the SRs, it is interesting that  
542 the LT and LE core motifs also form the basis of the 20 bp rabbit SR, differing by five  
543 interrupting nucleotides, GTACG. It is noteworthy that these can be converted by C>T and  
544 A>G transitions to form a LT motif 2 + LE motif 20 bp consensus sequence block (Fig. 3B). It  
545 is likely that such transition mutations can occur repeatedly during evolution, while the G>C  
546 transversion, separating the LT motif 2 and 3 part of the rabbit motif from the rest (Fig. 3B), is  
547 unlikely as a recurrent point mutation. Alternatively, it is possible that the LE motif and LT  
548 motif 2 are recombinants of two ancestral repeat types, which are present (in modified state) in  
549 rabbits. Overall, the remarkable conservation of the core LT-type SR repeats in hares (*Lepus*)  
550 is intriguing (Fig. 3C). This suggests that their origin predates the separation of these species  
551 and highlights the unique situation of *L. europaeus* SRs, whose core sequence is not shared by  
552 the other *Lepus* species, including its close relative, the cape hare (*L. capensis*) (Ferreira et al.,

553 2021). Nevertheless, the identical length of the SRs core sequences (10 bp) and their similarity  
554 in all *Lepus* species (Fig. 3B, only a single nucleotide difference separates the *L. europaeus*  
555 core sequence from each of two SR cores found in other *Lepus* species) can be only explained  
556 by a common ancestry or mechanism of origin among all species in the genus. The sequence  
557 homogeny of the SR runs within a species is quite likely maintained by constant contraction  
558 and reamplification of the repeat motifs, resulting in their rapid drift.

559 Three more observations are worth mentioning concerning the SRs region. (a) In all  
560 species and independently of the SR core, the repeats located at the 3' end and in the center of  
561 the SR are the least conserved (Fig. 3C). Similar asymmetry is known to result from the  
562 unidirectional gene conversion associated with some mitochondrial mobile elements or with  
563 mating type switching in yeast (Richard et al., 2008; Stoddard, 2014), which includes a double-  
564 strand break followed by erosion of the ends that generates nucleotide variation after repair. (b)  
565 Amongst the 48 core sequence variants in LE-type and LT-type of repeats, no LT-type core  
566 variant could be observed in species with LE-type core repeats and *vice versa*, even though the  
567 SR core sequences show very strong homology,. This is suprising, as it suggests that a unique  
568 SR motif amplification event has taken place specifically in the brown hare evolutionary  
569 lineage. Sequencing mitochondrial SRs from more *Lepus* species and isolates would be needed  
570 to allow drawing further hypotheses from this observation. (c) One species, *L. yarkandensis*,  
571 presents a unique SR motive constituted of LT motif 1 + LT motif 2 + a two-nucleotide (TT or  
572 TC) spacer. Similar short footprint signatures have been associated to the excision of some  
573 transposons (Skipper et al., 2013) as well as non-homologous end joining repair mechanisms  
574 (Yant and Kay, 2003).

575 The biological significance or functional roles of the SRs remain enigmatic. It is plausible  
576 that the SRs could form secondary structures when being single-stranded, which could promote  
577 priming by MTRPOL, as occurs during the initiation of L-strand replication at OriL (Fuste et

578 al., 2010). Interestingly, hairpin formation is predicted to be stronger for the H-strand (Fig. 4),  
579 which – following the OriL priming mechanism – would indicate a L-strand origin. Similarly,  
580 the reported variation in the SR copy number (Casane et al., 1997; Dufresne et al., 1996) and  
581 their proneness to mutate through strand-slippage (Pfeuty et al., 2001), are counterintuitive  
582 features for important genomic functions such as the regulatory control of DNA-replication.

583 Like the short repeats, the long repeats contain a core (20 bp) sequence segment (Fig.  
584 4A), which is conserved in all lagomorphs (Casane et al., 1997; Dufresne et al., 1996).  
585 Coincidentally, this is the same length as the duplex SR motifs, although the sequence bears no  
586 resemblance to them and has no obvious sequence features such as palindromes. If it represents  
587 a regulatory element capable of binding protein factors such as TFAM, as proposed (Dufresne  
588 et al., 1996), the existence of multiple such sites on mtDNA is somewhat reminiscent of the  
589 initiator titration model of bacterial replication initiation (Hansen et al., 1991; Ogawa et al.,  
590 2002). In *E. coli* for example, the initiation factor DnaA is titrated away from the replication  
591 origin *oriC* binding sites by providing competing recognition sites dispersed elsewhere on the  
592 genome. Replication of the genome causes the duplication of these sites, binding more DnaA  
593 and thereby reducing its availability to initiate new rounds of replication. The obvious problem  
594 for this mechanism to operate in the regulation of hare mtDNA replication initiation is the  
595 variability in the copy number of the LRs (Fig. 4B, C), where having three to nine copies of  
596 competitive binding sites per genome would effectively negate any accurate regulation,  
597 although this could be balanced by a difference in mtDNA copy number and TFAM molecules.  
598 Unlike the dispersed DnaA sites, the LRs are clustered together and situated so that their  
599 doubling would occur only at the end of the replication. Thus, they are unlikely to prevent re-  
600 initiation before termination. Still, they could control replication by titrating TFAM in function  
601 of the number of mtDNA copies (and therefore SRs repeats) already present in the  
602 mitochondria. However, as with the SRs, if these elements have such an important role in

603 genome regulation, why would they be present in some evolutionary lineages but not in others?  
604 For example, rodents (Rodentia) – a sister order of Lagomorpha – as well as most other  
605 mammals do not have such repetitive sequence arrays on their mtDNAs.

606 The species difference in the LR length distribution between brown hare and mountain  
607 hare (Fig. 6A) is interesting, as it tells that their variation between individuals is not random.  
608 In fact, similar differences in mtDNA tandem repeat lengths are known from other closely  
609 related species (Hernández et al., 2004; Lunt et al., 1998; Mundy and Helbig, 2004; Omote et  
610 al., 2013). As the species difference is fundamentally a genetic difference, one might expect  
611 that the LR copy number would then also be governed by the nuclear genome. However, the  
612 fact that LR copy number can vary between the tissues of the same individual in rabbits and  
613 that certain tissues are more prone to have longer arrays than others (Casane et al., 1997),  
614 suggests a more complicated mechanism, where the nuclear genetic background is only one  
615 variable. As different tissues have different metabolic requirements, manifesting as differences  
616 in mitochondrial activity and mtDNA maintenance strategies (Herbers et al., 2019), it is  
617 plausible that the internal environment in mitochondria also indirectly effects the LR-region  
618 length through factors such as temperature, oxidative stress and replication/repair protein  
619 availability. In fact, we previously noted that mountain hare fibroblasts maintain higher  
620 mitochondrial membrane potential than brown hare cells (Gaertner et al., 2023) and have also  
621 other differences in their oxidative metabolism. In this case, the energetic state of the cells can  
622 be influenced by variables such as metabolic adaptations, tissue type, ageing and stress, causing  
623 also the observed variation in the LR copy numbers. If the LRs are generated or lost as a result  
624 of strand-slippage during replication, as seems to be the case with the SRs (Pfeuty et al., 2001),  
625 changes in the mitochondrial internal environment could increase the probability of such  
626 events, resulting in LR copy number change. Similarly, factors such as mitochondrial  
627 temperature (Chretien et al., 2018), could favor certain repeat lengths over others and stabilize

628 the LR copy number variation. Overall, it is curious that the SR core sequence shows less  
629 conservation (Fig. 3C) and the LR arrays show more length variation in brown hares than in  
630 mountain hares.

631 Interestingly, our experiments with the cybrids showed mixed results (Fig. 6B). First, the  
632 effect on LR copy number was asymmetric and only presented by cells with mountain hare  
633 nucleus, regardless of the mtDNA origin. Also, the change in the LR array length in was not  
634 consistent. While in one cybrid line the brown hare mtDNA lost a LR copy, obtaining the  
635 common copy number of five for a mountain hare mtDNA, in another the copy number was  
636 increased to seven, a LR array length not presented by any of the parental cell lines. In contrast,  
637 shorter than parental mountain hare LR arrays appeared in the control cybrid with LT6 nucleus  
638 and LT4 mtDNA. Although the sample size is small, it is interesting that the observed effect is  
639 consistent with the frequent introgression of mountain hare mtDNA into brown hare population  
640 but not *vice versa* (Levanen et al., 2018a; Thulin and Tegelström, 2002), suggesting that  
641 mtDNA maintenance is perhaps more permissive or resilient in cells with brown hare nucleus.  
642 Alternatively, the mountain hare cells could be more sensitive to the stress caused by the  
643 process of generating cybrids. In all cases, the new haplotype appeared in heteroplasmy with  
644 the old, showing that they arise as a spontaneous rearrangement of the existing LRs and  
645 increase in abundance through drift or selection.

646

## 647 **5 Conclusions**

648 Although being a decades old discovery, repeat elements on the non-coding region remain an  
649 overlooked feature of Lagomorph mitochondrial genomes. We recognized conserved 10 bp  
650 core sequences in the short repeats, with two different motifs present in most hare species and  
651 one in brown hares. Interestingly, SR core element copies showed sequence variation in brown  
652 hares, but not in other hare species. We also found that while the long repeat arrays are dynamic,

653 their presentation is not random, as the different species show different length distributions.  
654 Considering their absence from most mammalian species, as well as the lack of a clear  
655 biological role in mtDNA maintenance and expression, it is likely that both repeat types have  
656 arisen and are maintained as selfish genetic elements. It is unlikely that these elements are  
657 mobile, and they are certainly too small to encode functional gene products, indicating that  
658 their reproduction has to be related to the mechanism of mtDNA replication. Why such  
659 elements have not been lost from the otherwise compact mitochondrial genomes and how  
660 species-specific copy number variation is maintained, warrants further research.

## 661 **6 Acknowledgements**

662 Dr Seppo Helisalmi and Dr Joose Raivo from the Gene diagnostics lab at Genome Center of  
663 Eastern Finland are thanked for providing the mtDNA sequencing service. We thank Ms Anita  
664 Kervinen for her valuable laboratory assistance in preparing the mtDNA PCRs for the  
665 sequencing. Funding: This research was supported by the Academy of Finland (grant no.  
666 329264 for ED and JLOP). The funder had no role in study design; sampling, analysis and  
667 interpretation of data; in the writing of the report; and in the decision to submit the article for  
668 publication.

## 669 **7 References**

670 Adams, K.L., Palmer, J.D., 2003. Evolution of mitochondrial gene content: gene  
671 loss and transfer to the nucleus. *Mol Phylogenet Evol* 29, 380-395.  
672 Bayona-Bafaluy, M.P., Manfredi, G., Moraes, C.T., 2003. A chemical enucleation  
673 method for the transfer of mitochondrial DNA to  $\rho^0$  cells -: art. no. e98.  
674 *Nucleic Acids Research* 31.  
675 Birney, E., Clamp, M., Durbin, R., 2004. GeneWise and Genomewise. *Genome*  
676 *Res* 14, 988-995.  
677 Boore, J.L., 1999. Animal mitochondrial genomes. *Nucleic Acids Res* 27, 1767-  
678 1780.  
679 Casane, D., Dennebouy, N., de Rochambeau, H., Mounolou, J.C., Monnerot, M.,  
680 1997. Nonneutral evolution of tandem repeats in the mitochondrial DNA  
681 control region of lagomorphs. *Mol Biol Evol* 14, 779-789.

682 Chen, S.F., Zhou, Y.Q., Chen, Y.R., Gu, J., 2018. fastp: an ultra-fast all-in-one  
683 FASTQ preprocessor. *Bioinformatics* 34, 884-890.

684 Chretien, D., Benit, P., Ha, H.H., Keipert, S., El-Khoury, R., Chang, Y.T.,  
685 Jastroch, M., Jacobs, H.T., Rustin, P., Rak, M., 2018. Mitochondria are  
686 physiologically maintained at close to 50 degrees C. *PLoS Biol* 16,  
687 e2003992.

688 Dufresne, C., Mignotte, F., Gueride, M., 1996. The presence of tandem repeats  
689 and the initiation of replication in rabbit mitochondrial DNA. *Eur J Biochem*  
690 235, 593-600.

691 Ennafaa, H., Monnerot, M., Elgaaied, A., Mounolou, J.C., 1987. Rabbit  
692 Mitochondrial-DNA - Preliminary Comparison between Some Domestic  
693 and Wild Animals. *Genet Sel Evol* 19, 279-288.

694 Ferreira, M.S., Jones, M.R., Callahan, C.M., Farelo, L., Tolesa, Z., Suchentrunk,  
695 F., Boursot, P., Mills, L.S., Alves, P.C., Good, J.M., Melo-Ferreira, J., 2021.  
696 The Legacy of Recurrent Introgression during the Radiation of Hares. *Syst  
697 Biol* 70, 593-607.

698 Fontanesi, L., Tazzoli, M., Pecchioli, E., Hauffe, H.C., Robinson, T.J., Russo, V.,  
699 2008. Sexing European rabbits (*Oryctolagus cuniculus*), European brown  
700 hares (*Lepus europaeus*) and mountain hares (*Lepus timidus*) with ZFX and  
701 ZFY loci. *Mol Ecol Resour* 8, 1294-1296.

702 Formenti, G., Rhie, A., Balacco, J., Haase, B., Mountcastle, J., Fedrigo, O.,  
703 Brown, S., Capodiferro, M.R., Al-Ajli, F.O., Ambrosini, R., Houde, P.,  
704 Koren, S., Oliver, K., Smith, M., Skelton, J., Betteridge, E., Dolucan, J.,  
705 Corton, C., Bista, I., Torrance, J., Tracey, A., Wood, J., Uliano-Silva, M.,  
706 Howe, K., McCarthy, S., Winkler, S., Kwak, W., Korlach, J.,  
707 Fungtammasan, A., Fordham, D., Costa, V., Mayes, S., Chiara, M., Horner,  
708 D.S., Myers, E., Durbin, R., Achilli, A., Braun, E.L., Phillip, A.M., Jarvis,  
709 E.D., Vertebrate Genomes Project, C., 2021. Complete vertebrate  
710 mitogenomes reveal widespread repeats and gene duplications. *Genome  
711 Biol* 22, 120.

712 Fuste, J.M., Wanrooij, S., Jemt, E., Granycome, C.E., Cluett, T.J., Shi, Y.,  
713 Atanassova, N., Holt, I.J., Gustafsson, C.M., Falkenberg, M., 2010.  
714 Mitochondrial RNA polymerase is needed for activation of the origin of  
715 light-strand DNA replication. *Mol Cell* 37, 67-78.

716 Gaertner, K., Michell, C., Tapanainen, R., Goffart, S., Saari, S., Soininmaki, M.,  
717 Dufour, E., Pohjoismaki, J.L.O., 2023. Molecular phenotyping uncovers  
718 differences in basic housekeeping functions among closely related species  
719 of hares (*Lepus* spp., Lagomorpha: Leporidae). *Mol Ecol* 32, 4097-4117.

720 Gertz, E.M., Yu, Y.K., Agarwala, R., Schaffer, A.A., Altschul, S.F., 2006.  
721 Composition-based statistics and translated nucleotide searches: improving  
722 the TBLASTN module of BLAST. *BMC Biol* 4, 41.

723 Hansen, F.G., Christensen, B.B., Atlung, T., 1991. The initiator titration model:  
724 computer simulation of chromosome and minichromosome control. *Res  
725 Microbiol* 142, 161-167.

726 Herbers, E., Kekalainen, N.J., Hangas, A., Pohjoismaki, J.L., Goffart, S., 2019.  
727 Tissue specific differences in mitochondrial DNA maintenance and  
728 expression. *Mitochondrion* 44, 85-92.

729 Hernández, M.A., Campos, F., Gutiérrez-Corcher, F., Amezcuia, A., 2004.  
730 Identification of Lanius species and subspecies using tandem repeats in the  
731 mitochondrial DNA control region. *Ibis* 146, 227-230.

732 Hoelzel, A.R., 1993. Evolution by DNA turnover in the control region of  
733 vertebrate mitochondrial DNA. *Current Opinions in Genetics and  
734 Development* 3, 891-895.

735 Jemt, E., Persson, O., Shi, Y., Mehmedovic, M., Uhler, J.P., Davila Lopez, M.,  
736 Freyer, C., Gustafsson, C.M., Samuelsson, T., Falkenberg, M., 2015.  
737 Regulation of DNA replication at the end of the mitochondrial D-loop  
738 involves the helicase TWINKLE and a conserved sequence element. *Nucleic  
739 Acids Res* 43, 9262-9275.

740 Juhling, F., Putz, J., Bernt, M., Donath, A., Middendorf, M., Florentz, C., Stadler,  
741 P.F., 2012. Improved systematic tRNA gene annotation allows new insights  
742 into the evolution of mitochondrial tRNA structures and into the  
743 mechanisms of mitochondrial genome rearrangements. *Nucleic Acids Res*  
744 40, 2833-2845.

745 Kasamatsu, H., Robberson, D.L., Vinograd, J., 1971. A novel closed-circular  
746 mitochondrial DNA with properties of a replicating intermediate. *Proc Natl  
747 Acad Sci U S A* 68, 2252-2257.

748 Kikin, O., D'Antonio, L., Bagga, P.S., 2006. QGRS Mapper: a web-based server  
749 for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res*  
750 34, W676-682.

751 Kobayashi, T., Ganley, A.R., 2005. Recombination regulation by transcription-  
752 induced cohesin dissociation in rDNA repeats. *Science* 309, 1581-1584.

753 Kuhl, I., Miranda, M., Posse, V., Milenkovic, D., Mourier, A., Siira, S.J.,  
754 Bonekamp, N.A., Neumann, U., Filipovska, A., Polosa, P.L., Gustafsson,  
755 C.M., Larsson, N.G., 2016. POLRMT regulates the switch between  
756 replication primer formation and gene expression of mammalian mtDNA.  
757 *Sci Adv* 2, e1600963.

758 Levanen, R., Kunnasranta, M., Pohjoismaki, J., 2018a. Mitochondrial DNA  
759 introgression at the northern edge of the brown hare (*Lepus europaeus*)  
760 range. *Annales Zoologici Fennici* 55, 15-24.

761 Levanen, R., Thulin, C.G., Spong, G., Pohjoismaki, J.L.O., 2018b. Widespread  
762 introgression of mountain hare genes into Fennoscandian brown hare  
763 populations. *Plos One* 13.

764 Li, D., Liu, C.M., Luo, R., Sadakane, K., Lam, T.W., 2015. MEGAHIT: an ultra-  
765 fast single-node solution for large and complex metagenomics assembly via  
766 succinct de Bruijn graph. *Bioinformatics* 31, 1674-1676.

767 Lunt, D.H., Whipple, L.E., Hyman, B.C., 1998. Mitochondrial DNA variable  
768 number tandem repeats (VNTRs): utility and problems in molecular  
769 ecology. *Mol Ecol* 7, 1441-1455.

770 Melo-Ferreira, J., Boursot, P., Suchentrunk, F., Ferrand, N., Alves, P.C., 2005.  
771 Invasion from the cold past: extensive introgression of mountain hare (*Lepus*  
772 *timidus*) mitochondrial DNA into three other hare species in northern Iberia.  
773 *Mol Ecol* 14, 2459-2464.

774 Melo-Ferreira, J., Vilela, J., Fonseca, M.M., da Fonseca, R.R., Boursot, P., Alves,  
775 P.C., 2014. The elusive nature of adaptive mitochondrial DNA evolution of  
776 an arctic lineage prone to frequent introgression. *Genome Biol Evol* 6, 886-  
777 896.

778 Meng, G., Li, Y., Yang, C., Liu, S., 2019. MitoZ: a toolkit for animal  
779 mitochondrial genome assembly, annotation and visualization. *Nucleic  
780 Acids Res* 47, e63.

781 Mundy, N.I., Helbig, A.J., 2004. Origin and evolution of tandem repeats in the  
782 mitochondrial DNA control region of shrikes (*Lanius* spp.). *J Mol Evol* 59,  
783 250-257.

784 Ogawa, T., Yamada, Y., Kuroda, T., Kishi, T., Moriya, S., 2002. The datA locus  
785 predominantly contributes to the initiator titration mechanism in the control  
786 of replication initiation in *Escherichia coli*. *Mol Microbiol* 44, 1367-1375.

787 Omote, K., Nishida, C., Dick, M.H., Masuda, R., 2013. Limited phylogenetic  
788 distribution of a long tandem-repeat cluster in the mitochondrial control  
789 region in Bubo (Aves, Strigidae) and cluster variation in Blakiston's fish owl  
790 (*Bubo blakistoni*). *Mol Phylogenet Evol* 66, 889-897.

791 Persson, O., Muthukumar, Y., Basu, S., Jenninger, L., Uhler, J.P., Berglund, A.K.,  
792 McFarland, R., Taylor, R.W., Gustafsson, C.M., Larsson, E., Falkenberg, M.,  
793 2019. Copy-choice recombination during mitochondrial L-strand synthesis  
794 causes DNA deletions. *Nat Commun* 10, 759.

795 Pfeuty, A., Gueride, M., Lecellier, G., 2001. Expansion/contraction of  
796 mammalian mitochondrial DNA repeats in *Escherichia coli* mimics the  
797 mitochondrial heteroplasmy. *J Mol Biol* 314, 709-716.

798 Pham, X.H., Farge, G., Shi, Y., Gaspari, M., Gustafsson, C.M., Falkenberg, M.,  
799 2006. Conserved sequence box II directs transcription termination and  
800 primer formation in mitochondria. *J Biol Chem* 281, 24647-24652.

801 Phillips, A.F., Millet, A.R., Tigano, M., Dubois, S.M., Crimmins, H., Babin, L.,  
802 Charpentier, M., Piganeau, M., Brunet, E., Sfeir, A., 2017. Single-Molecule  
803 Analysis of mtDNA Replication Uncovers the Basis of the Common  
804 Deletion. *Mol Cell* 65, 527-538 e526.

805 Pohjoismaki, J.L.O., Forslund, J.M.E., Goffart, S., Torregrosa-Munumer, R.,  
806 Wanrooij, S., 2018. Known Unknowns of Mammalian Mitochondrial DNA  
807 Maintenance. *Bioessays* 40, e1800102.

808 Pohjoismaki, J.L.O., Michell, C., Levanen, R., Smith, S., 2021. Hybridization  
809 with mountain hares increases the functional allelic repertoire in brown  
810 hares. *Sci Rep* 11, 15771.

811 Richard, G.F., Kerrest, A., Dujon, B., 2008. Comparative genomics and molecular  
812 dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev* 72, 686-  
813 727.

814 Savolainen, P., Arvestad, L., Lundeberg, J., 2000. mtDNA tandem repeats in  
815 domestic dogs and wolves: mutation mechanism studied by analysis of the  
816 sequence of imperfect repeats. *Mol Biol Evol* 17, 474-488.

817 Skipper, K.A., Andersen, P.R., Sharma, N., Mikkelsen, J.G., 2013. DNA  
818 transposon-based gene vehicles - scenes from an evolutionary drive. *J  
819 Biomed Sci* 20, 92.

820 Solignac, M., Monnerot, M., Mounolou, J.C., 1986. Concerted Evolution of  
821 Sequence Repeats in *Drosophila* Mitochondrial-DNA. *J Mol Evol* 24, 53-  
822 60.

823 Spinelli, J.B., Haigis, M.C., 2018. The multifaceted contributions of  
824 mitochondria to cellular metabolism. *Nat Cell Biol* 20, 745-754.

825 Stoddard, B.L., 2014. Homing endonucleases from mobile group I introns:  
826 discovery to genome engineering. *Mob DNA* 5, 7.

827 Thakur, J., Packiaraj, J., Henikoff, S., 2021. Sequence, Chromatin and Evolution  
828 of Satellite DNA. *Int J Mol Sci* 22.

829 Thulin, C.-G., Tegelström, H., 2002. Biased geographical distribution of  
830 mitochondrial DNA that passed the species barrier from mountain hares to  
831 brown hares (genus *Lepus*): an effect of genetic incompatibility and mating  
832 behaviour? *J Zool* 258, 299-306.

833 Uliano-Silva, M., Ferreira, J., Krasheninnikova, K., Darwin Tree of Life, C.,  
834 Formenti, G., Abueg, L., Torrance, J., Myers, E.W., Durbin, R., Blaxter, M.,  
835 McCarthy, S.A., 2023. MitoHiFi: a python pipeline for mitochondrial  
836 genome assembly from PacBio high fidelity reads. *BMC Bioinformatics* 24,  
837 288.

838 Wang, T.Y., Tzeng, C.S., Teng, H.Y., Chang, T., 2007. Phylogeography and  
839 identification of a 187-bp-long duplication within the mitochondrial control  
840 region of  
841 (Teleostei: Balitoridae). *Zool Stud* 46, 569-582.

842 Wang, X., Liu, N., Zhang, H., Yang, X.J., Huang, Y., Lei, F., 2015. Extreme  
843 variation in patterns of tandem repeats in mitochondrial control region of  
844 yellow-browed tits (*Sylviparus modestus*, Paridae). *Sci Rep* 5, 13227.

845 Wanrooij, S., Fuste, J.M., Farge, G., Shi, Y., Gustafsson, C.M., Falkenberg, M.,  
846 2008. Human mitochondrial RNA polymerase primes lagging-strand DNA  
847 synthesis in vitro. *Proc Natl Acad Sci U S A* 105, 11122-11127.

848 Wheeler, T.J., Eddy, S.R., 2013. nhmmer: DNA homology search with profile  
849 HMMs. *Bioinformatics* 29, 2487-2489.

850 Yant, S.R., Kay, M.A., 2003. Nonhomologous-end-joining factors regulate DNA  
851 repair fidelity during Sleeping Beauty element transposition in mammalian  
852 cells. *Mol Cell Biol* 23, 8505-8518.

853