High quality genome assembly of the brown hare (*Lepus europaeus*) with chromosome-level scaffolding

Craig Michell[1,2], Joanna Collins[3], Pia K. Laine[4], Zsófia Fekete[1,5], Riikka Tapanainen[1], Jonathan M. D. Wood[3], Steffi Goffart[1] and Jaakko L. O. Pohjoismäki*[1]

[1]Department of Environmental and Biological Sciences, University of Eastern Finland – Joensuu, Finland
[2]Red Sea Research Center, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, – Thuwal, Kingdom of Saudi Arabia
[3]Tree of Life, Wellcome Sanger Institute, Hinxton, Cambridge, UK.
[4]Insitute of Biotechnology, University of Helsinki, Finland.
[5]Hungarian University of Agriculture and Life Sciences, Institute of Genetics and Biotechnology, Gödöllő, Hungary

*Corresponding author
Correspondence: jaakko.pohjoismaki@uef.fi

Abstract

We present here a high-quality genome assembly of the brown hare (*Lepus europaeus* Pallas), based on a fibroblast cell line of a male specimen from Liperi, Eastern Finland. This brown hare genome represents the first Finnish contribution to the European Reference Genome Atlas pilot effort to generate reference genomes for European biodiversity.

The genome was assembled using 25X PacBio HiFi sequencing data and scaffolded utilizing a Hi-C chromosome structure capture approach. After manual curation, the assembled genome length was 2,930,972,003 bp with N50 scaffold of 125.8 Mb. 93.16% of the assembly could be assigned to 25 identified chromosomes (23 autosomes plus X and Y), matching the published karyotype. The chromosomes were numbered according to size. The genome has a high degree of completeness based on the BUSCO score (mammalia_odb10 database), Complete: 96.1% [Single copy: 93.1%, Duplicated: 3.0%], Fragmented 0.8%, and Missing 2.9%. The mitochondrial genome of the cell line was sequenced and assembled separately. The final annotated genome has 30,833 genes of which 21,467 code for a polypeptide.

The brown hare genome is particularly interesting as this species readily hybridizes with the mountain hare (*Lepus timidus* L.) at the species contact zone in northern Eurasia, producing fertile offspring and resulting in gene flow between the two species. In addition to providing a useful comparison for population studies, the genome can offer insight into the chromosomal evolution among Glires in general and Lagomorpha in particular. The chromosomal assembly of the genome also demonstrates that the cell line has not acquired karyotypic changes during culture.

Keywords: Lagomorpha, reference genome, chromosome structure, phylogeny, fibroblast cell line

Introduction

The brown hare (*Lepus europaeus* Pallas), also known as the European hare, is a widespread species in the western parts of Eurasia (Bock, 2020). Besides its native range, the brown hare has been introduced to numerous regions, including the British Isles, the Falkland Islands, Canada, South America, Australia and New Zealand (Jaksic, 2023; Petrovan, 2013). In many places the brown hare is regarded as an invasive species and a threat to the local ecosystems (Stott, 2003) or native species (Reid, 2011).

The brown hare is a steppe-adapted species living in open grasslands and avoiding forested regions. Its colonization history in Europe is complex. The species has had glacial refugia during the Pleistocene in the Italian peninsula, the Balkans and Asia Minor (Fickel et al., 2008), with an interesting pre-glacial diversity hotspot in the archipelago of Greece (Minoudi et al., 2018). After the ice age, the species has been expanding its range both naturally as well as facilitated by human-caused changes in the landscape, especially through the expansion of agricultural lands and pastures. As mentioned above, the brown hare has been also frequently introduced by humans to new areas from antiquity to the present day (Petrovan, 2013), resulting in potential mixing of ancestral populations. This is true also for Finland, where the species has arrived naturally from the southeast through the Karelian isthmus in the late 19th century (Ognev, 1940; Siivonen, 1972; Wadén, 1910) but also through local introductions to the southwestern parts of the country in the 1910s (Rikala, 1925; Salonius, 1924; Suomalainen, 1922; Wadén, 1910).

Curiously, Linné was unaware of the existence of the brown hare, despite having described the mountain hare (*Lepus timidus* L.) and the South African Cape hare (*Lepus capensis* L.). In fact, the brown hare is not native to Linné's home country Sweden but was introduced there only a century after his death (Thulin, 2003). Curiously, the species author Peter Simon Pallas never

published a formal description of the brown hare. Instead, the species was authority has been attributed to him because of inclusion of the name *Lepus europaeus* into his identification table of hare species of the world (Pallas, 1778: 30) (Figure 1A). Consequently, there is also no information of the type locality or any other information of the species in the original publication. The type locality of the brown hare has been retrospectively assigned – without too convincing arguments – to Burgundy (France) or Poland (Holden, 2005). Identifying Poland as the type locality was done by the Russian zoologist Sergey Ognev, and is based on Pallas mentioning in the 1778 text the hybrid forms of *Lepus variabilis* (synonym of *L. timidus*) and *L. europaeus* from Poland and Lithuania, as well as citing later works of Pallas for the species distribution (Ognev, 1940). In fact, Ognev (1940: 141) boldly states that "we consider southwestern Poland as a typical terrain for the nominal subspecies; there are hares with the characteristics of the basic form, expressed with full clarity" (translation from Russian).

While being widespread in the temperate regions of Western and Central Europe to the Caspian steppes in historical times, the brown hare began a northward expansion in the 19th century. In the Fennoscandian region, the species reached the St Petersburg area by 1820s (Ognev, 1940), arriving in Finland at the turn of the century (Salonius, 1924; Siivonen, 1972; Thulin, 2003; Wadén, 1910). By 1930, the species had become established throughout southern Finland, with the range expansion stagnating southwest of the N64°-N62° -line until the 1990s. Quite likely benefitting from the ongoing anthropogenic climate change, brown hares range expansion has intensified during the last three decades, currently reaching the polar circle (Levänen et al., 2018a) (Figure 1B), as well as higher altitudes in the Alps (Schai-Braun et al., 2023). In Finland, the expansion seems to be limited with 150 days of annual snow cover (Levänen et al., 2018a). The range expansion together with the past introductions have brought the brown hare increasingly in contact with the mountain hare, especially at the northern edges of the species distribution, constituting a threat especially to the more temperate climate adapted populations and subspecies of the mountain hare (Reid, 2011; Reid and Montgomery, 2007; Thulin, 2003; Thulin et al., 2021). Besides competing, brown hares also hybridize with mountain hares, producing fertile offspring and resulting in gene flow between the species. This gene flow is biased towards the brown hare (Ferreira et al., 2021; Levänen et al., 2018a; Levänen et al., 2018b; Thulin and Tegelström, 2002), which obtains genetic variation from mountain hares, some of which might have adaptive significance (Pohjoismäki et al., 2021). In contrast to the brown hare's expansion towards north, many of the Central European brown hare populations are contracting in range and numbers, likely driven by changes in agricultural practices and land use (Schai-Braun et al., 2013; Smith et al., 2005).

Overall, the brown hare's features as an invasive species outside Europe and expansive species in the north, its hybridization tendency and complex population ancestries make the species highly interesting not only to study population genetics and adaptation mechanisms, but also for the understanding the genomic makeup of species boundaries (Gaertner et al., 2022). These studies would be greatly facilitated by the availability of high quality, chromosomally scaffolded reference genomes (Blaxter et al., 2022).

The last decade has witnessed a boom in genome sequencing technologies, leading to many sequencing initiatives, such as the Earth BioGenome Project (EBP), European Reference Genome Atlas (ERGA) and the Darwin Tree of Life (DToL), aiming to produce high quality reference genomes for a variety of organisms (Blaxter et al., 2022). This has been enabled by the

technological development of high-throughput sequencing of long molecules. The read length N50's > 19 kbp, allows a high contiguity compared to the genome assemblies released in the early 2010's that were based on short read sequencing technologies. Additionally, Hi-C sequencing, a method measuring the contact frequency between all pairs of loci in the genome, has allowed producing pretext maps of chromosomes (Ghurye et al., 2019), enabling chromosome level assemblies with a precision that has been previously achieved only with model organisms (Lawniczak et al., 2022).

In the presented work, we apply the PacBio HiFi long-range sequencing technology with Hi-C sequencing to produce a high-quality reference genome of the brown hare (*Lepus europaeus*). While there is no previous genome assembly for the species, a mountain hare assembly (GCA_009760805) exists, representing the Irish subspecies *L. t. hibernicus* and is based on a female specimen, thus lacking the Y-chromosome. This assembly is also a so-called "pseudoreference", assembled to chromosomes using the rabbit (*Oryctolagus cuniculus* L.) reference genome (Marques et al., 2020). The original rabbit reference genome had been established using whole-genome shotgun sequencing of females representing the inbred Thorbecke New Zealand White line and was quite fragmented (Carneiro et al., 2014), updated only recently with long-read sequencing data (Bai et al., 2021). Notably, the diploid chromosome number in domestic rabbits is 2n = 44 (Korstanje et al., 1999), whereas it is 48 in brown hare and mountain hare (Gustavsson, 1972).

To obtain a reference genome conforming to current standards, we used a fibroblast cell line (LE1) of a male specimen of brown hare (*Lepus europaeus*) from Liperi, Eastern Finland (Gaertner et al., 2022), as a source of high-molecular weight (HMW) DNA and fresh RNA. Compared to solid tissues, fibroblasts are optimal for Hi-C, as the method was originally developed for cells growing in a monolayer (Lieberman-Aiden et al., 2009). Now, over 240 years after the first mention of the name *Lepus europaeus* in the literature (Figure 1A), we have expanded the original seven-word description to encompass 2,930,972,003 genomic letters. These letters have been assembled in a highly complete (BUSCO mammalia_odb10 database, complete: 96.1%), and continuous (N50 scaffold: 125.8 Mb) manner with 93.16% of the sequence being placed in the identified 23 autosomes, and X and Y sex chromosomes. The presented genome assembly not only contains the identity of the brown hare as a species, but also provides a gateway to detailed knowledge of its biology and evolutionary history. Besides being the first Finnish contribution to the European Reference Genome Atlas (Mazzoni et al., 2023), the brown hare reference genome represents a continuing effort to map and understand our planets biodiversity.

Methods

Sampling and confirming of the species identity

A young male brown hare was sampled in October 2018 in Kuorinka, Liperi, Finland (62.6207 N, 29.4478 E, Figure 1B). In Finland, brown hares are highly dependent on similar man-made environments (Levänen et al., 2018a; Levänen et al., 2019), and also the collection location is agricultural area with a mosaic of fields, shrubs and mixed forest with a strong brown hare population and only occasional stray mountain hares. Geographically, the sampled population is part of the same distribution continuum as with species proposed type locality, through Russian

Karelia and the Baltic states to Poland. This is important, as is recommendable that the species reference genome would represent or be closely related to the type locality population (Lawniczak et al., 2022).

The sampling had minimal impact on the local brown hare population and no impact on the habitat. As brown hares are legal game animals in Finland and the hunting followed the regional hunting seasons and legislation (Metsästyslaki [Hunting law] 1993/615/5§), the sampling adheres to the ARRIVE guidelines and no ethical assessment for the sampling was required. The sampling also 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 species identity was confirmed at site based on the morphological features distinguishing the brown hare from the mountain hare, the only other hare species in northern Europe. Further analysis of the coding part of the LE1 genome and mitochondrial DNA haplotyping showed minimal ancestral admixture with mountain hares (Gaertner et al., 2022). In fact, unlike many other brown hares in the region, showing adaptive introgression of mountain hare specific UCP1 alleles (Pohjoismäki et al., 2021), the LE1 specimen is homozygous for the common ancestral brown hare allele, UCP02.

Generation and vouchering of the cell line

The fibroblast cell line (LE1) was isolated from the specimen as described earlier (Gaertner et al., 2022) and is deposited as cryopreserved living cells under voucher number ZFMK-TIS-69747 into the biobank of the Stiftung Leibniz-Institut zur Analyse des Biodiversitätswandels, Zoological Research Museum Alexander König (ZFMK), Bonn, Germany. Additional identifiers for the sample are mLepEur2 and ERGA specimen ID ERGA_FI_3610_002 (COPO portal, https://copo-project.org).

HMW DNA extraction and PacBio HiFi sequencing

A salting out-method was used to extract high molecular weight DNA from cells grown to confluency on a 10 cm cell culture dish. Briefly, the cells were detached from the dish using trypsin, followed by centrifugation and two washes with PBS. The salting out-method then followed the 10X Genomics “Salting Out Method for DNA extraction from cells” protocol (10x-Genomics, 2017). The high molecular weight DNA was quantified using a Qubit 3.0 (Thermo Fisher Scientific™, Malaysia), followed by qualification with an 0.8 % Agarose gel. The DNA was sent for the library preparation and sequencing at the DNA Sequencing and Genomics Laboratory, Institute of Biotechnology, University of Helsinki on the PacBio Sequel II. In brief, the DNA was quantified using Qubit and DNA fragment length was measured using Fragment Analyzer (HS large fragment kit). DNA was then sheared with Megaruptor 3 (Diagenode) (45/200/30) to produce DNA of length 24 kb. Buffer was then replaced with PacBio’s Elution buffer using AMPure beads. Repair and A-tailing, adapter ligation, nuclease treatment and cleanup with SMRTbell cleanup beads was then performed with the SMRTbell prep kit 3.0. Larger fragments (greater than 10 000 bp) were then purified with the BluePippin (Sage Science) (0.75% DF Marker S1 high pass 6-10kb v3) and DNA was further purified with PacBio AMPure beads and treated with DNA damage repair mix (PacBio) (37°C for 60 mins) and then purified with AMPure clean-up beads and eluted to 11 ul PacBio’s Elution buffer. Libraries were sequenced on the PacBio Sequel II at a concentration of 90 pM using PacBio’s instructions (provided by the SMRTlink software). Sequence data from two

flow cells was produced. Genomescope2 was used to visualize the HiFi sequencing results (Ranallo-Benavidez et al., 2020).

Mitochondrial DNA (mtDNA) sequencing

The mitochondrial genomes of our hare cell lines (Gaertner et al., 2022) have been sequenced for the purposes of another study using approximately 2 kb overlapping PCR-amplified fragments of mtDNA. The primers used to amplify the mountain 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

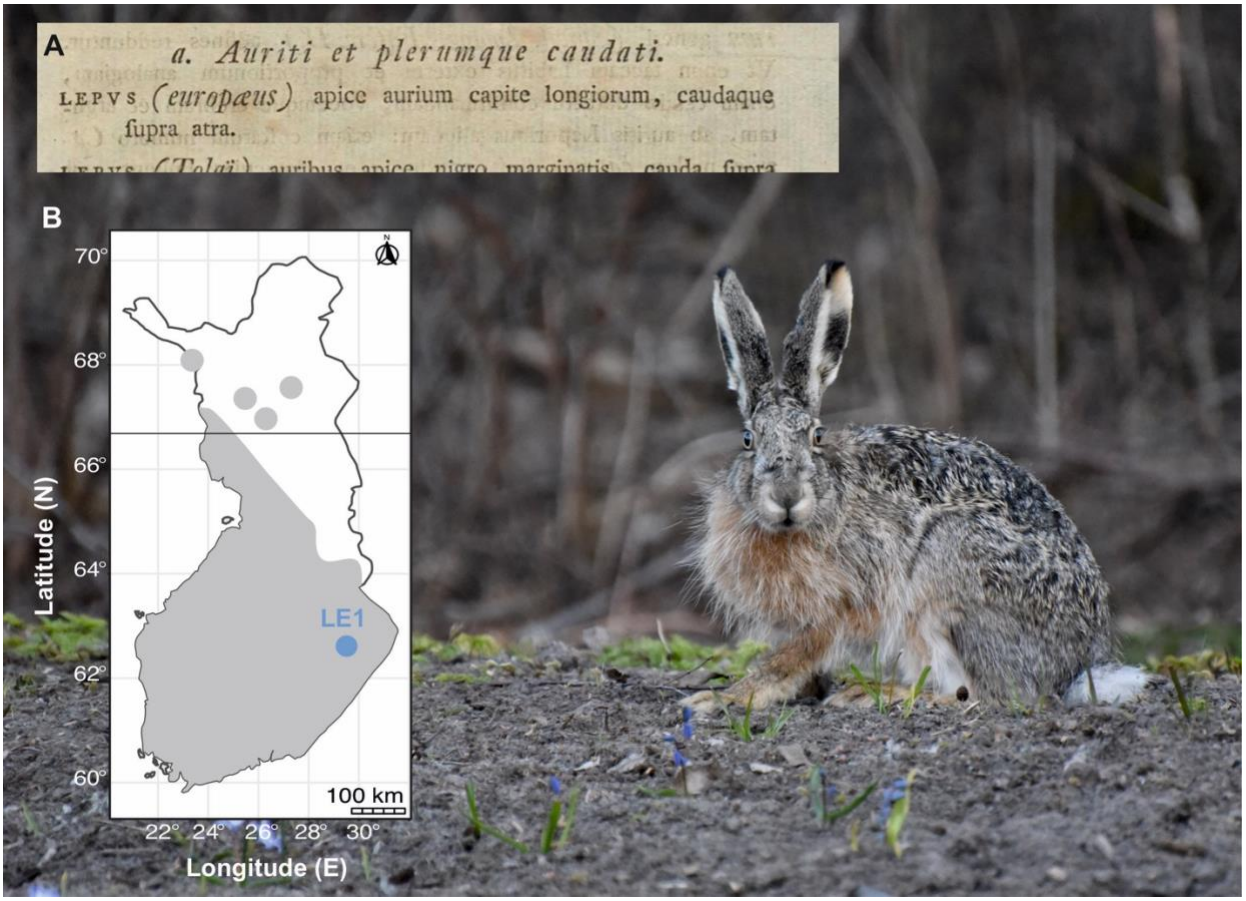

Figure 1. The brown hare. (A) An excerpt from the identification table by Pallas (1778: 30), which has been assigned as the species description for *Lepus europaeus*. "Tips of the ears longer than the head, tail black above." (translation from Latin). (B) The geographic location of the LE1 sample used in this study. The grey-shaded area shows the approximate current distribution of the brown hare in Finland, based on the data in the Finnish Biodiversity Information Facility (FinBIF, https://laji.fi/en). Individual records from the north are from settlemens or towns. The arctic circle (black line) runs along the 66.6 ° parallel. Background photo shows a typical Finnish male

brown hare. Photo taken in Joensuu, 20 km East from the LE1 sampling site by Dr. Mervi Kunnasranta.

Hi-C library preparation

Hi-C sequencing libraries were prepared following the protocol of (Belaghzal et al., 2017) with the following changes: 1.) To prepare a diverse sequencing library, we performed the Hi-C protocol in triplicates, 2.) Size fractionation was performed using the NEBNext Ultra II FS DNA module (New England Biolabs), 3.) Library preparation was performed using the NEBNext Ultra II library preparation kit for Illumina (New England Biolabs) and 4.) Library enrichment was performed using triplicate PCR reactions with six cycles of PCR. The PCR reactions were then purified using Ampure XP beads (Beckman Coulter Life Sciences) at a ratio of 0.9X. The final clean libraries were quantified using Qbit, followed by agarose gel electrophoresis to confirm the fragment size. The sequencing was performed on a single lane of the Illumina NovaSeq 6000 using the SP flowcell with paired-end chemistry 2 x 150bp.

Genome assembly

HiFiasm version 0.18.7 (Cheng et al., 2021) was used to assemble the 25X (rq >0.99) PacBio HiFi reads using the recommended arguments *-l3 –h1 and –h2 –primary* to integrate the Hi-C read data and produce a primary assembly. We then continued with the scaffolding of the primary assembly and annotation. To process the HiC data, we first mapped the Hi-C data to the primary genome assembly using BWA-mem version 0.7.17 (Li and Durbin, 2009) with the arguments *-SP*. The mapped reads were then parsed and filtered using pairtools version 1.0.2 (Open2C et al., 2023). To parse the aligned Hi-C reads, we used the options *–no-flip –dd-columns mapq –walks-policy mask*. The parsed pair data was then sorted and deduplicated using default arguments with pairtools. Finally, we selected unique-unique and unique-rescued pairs and split these into the pairs file and bam file for input in YaHS version 1.1 (Zhou et al., 2023). YAHS was run using the default parameters with the primary contig assembly and the filtered Hi-C bam file. HiFiasm operates as a haplotype-aware assembler, counting and assembling unique k-mers. Utilizing k-mer coverage information, the assembler identifies regions in the genome with expected heterozygous coverage and subsequently amalgamates them into a unified sequence. To refine the assembly further, Hi-C data is incorporated to elucidate the phasing of variants and ascertain connected genomic regions. Consequently, the assembler's output provides a primary assembly along with an alternative assembly, effectively separating the haplotypes by their distinct alleles within the genomic content (please see https://lh3.github.io/2021/04/17/concepts-in-phased-assemblies). Contiguity and general genome statistics were calculated using QUAST version 5.2.0 (Mikheenko et al., 2018). We assessed the completeness of the genome by calculating the number of complete single copy orthologs with BUSCO version 5.1.2 (Manni et al., 2021), using the mammalia_odb10 database as well as the more lineage specific glires_odb10 database.

Genome annotation

Repeat annotation of the genome was performed with EDTA version 2.1.0 (Ou et al., 2019), a *de novo* repeat identification pipeline. Using the repeat library produced by EDTA, we masked the scaffolded genome using RepeatMasker version 4.1.1 (Smit et al., 2013-2015). RNA-seq data from the same cell line (SRA accession number: SRR18740842) as well as RNA-seq data from other *L.*

timidus libraries (SRA accession number: SRR10020054, SRR10020055, SRR10020060, SRR10491719, SRR18740839, SRR18740840, and SRR18740841) was collected from the sequence read archive (SRA). The previously produced RNA-seq data (Gaertner et al., 2022) was trimmed using fastp version 0.23.2 (Chen et al., 2018) and mapped against the masked genome using HISAT2 version 2.1.0 (Kim et al., 2019) with the default parameters. Furthermore, we included the protein sequences of the assembled transcripts from this cell line as further evidence for the genome annotation. These lines of gene evidence were included in the annotation using BRAKER3 version 3.02 (Bruna et al., 2021). The original annotation has been further augmented by the NCBI Eukaryotic Genome Annotation Pipeline (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/). Telomeric sequences [AACCCT]n were identified using a Telomere Identification toolkit (tidk) version 0.2.31. The telomeric sequence copy number was then calculated in windows of 200kb for visualization using Circos (Krzywinski et al., 2009).

Mitochondrial DNA assembly and annotation

Mitochondrial DNA was assembled from the PCR-amplified and Illumina sequenced mtDNA using the MitoZ pipeline (Meng et al., 2019). After comparing the results of the pipeline's outputs with different kmer options, we selected the best assembly as final. Run options used in the final assembly were --clade Chordata –fastq_read_length 150, --requiring_taxa Chordata --genetic_code 2 --kmers_megahit 21 29 39 59 79 99 119 141. The tools invoked by the pipeline included fastp (Chen et al., 2018) for cleaning the raw data, MEGAHIT (Li et al., 2015) for assembly, after which sequences were filtered to ensure the correct taxa by HMMER (Wheeler and Eddy, 2013) and further filtered for completeness of protein-coding genes. Annotation steps were done using TBLASTN (Gertz et al., 2006), GeneWise (Birney et al., 2004) and MiTFi (Juhling et al., 2012). Final manual curation, the annotation of the non-coding region (NCR) as well as the illustration of the mitochondrial genome was done using 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.

Manual curation

The assembled and annotated genome was manually curated to further improve its quality based on Hi-C contact maps as described in (Howe et al., 2021), allowing to break or join erroneus scaffold assemblies and removing duplicated contigs.

Comparison with previous assembly

We performed a comparison of our scaffolded assembly to the current *L. timidus* genome assembly (NCBI Accession number: GCA_009760805). Mapping to the genome was performed using minimap2 version 2.21 (Li, 2018) with the arguments *-asm5*. A dot plot of the alignment was created using the R script pafCoordsDotPlotly.R (https://github.com/tpoorten/dotPlotly).

Results

The genome assemblies can be accessed via BioSample accession SAMN33984520 as well as BioProject accession PRJNA1009711 for the primary assembly and PRJNA1009710 for the

355 alternative assembly. The Final annotated assembly can be found under NCBI Reference
356 Sequence assembly accession GCF_033115175.1, which can be viewed also using the NCBI
357 Genome Data Viewer
358 (https://www.ncbi.nlm.nih.gov/genome/gdv/browser/genome/?id=GCF_033115175.1).
359
360 **Genome assembly**
361 The expected haploid genome size of the *L. europaeus* is similar to *L. timidus*, which has been
362 estimated to be 3.25 pg (3,178,500,000 bp) (Vinogradov, 1998), containing 23 autosomal
363 chromosomes and two sex chromosomes (Gustavsson, 1972). PacBio HiFi sequencing with two
364 flow cells resulted in 25 X coverage of the expected genome size. The sequence N50 of the HiFi
365 data was 19.97 kb. When k-mer size of 21 is used, the resulting genome size from the PacBio HiFi
366 data is 2.4 Gbp (Figure 2A). The Illumina sequencing of the Hi-C data produced 494,285,585 paired
367 reads, representing about 51 X coverage of the genome, with a duplication rate of 17 %. Assembly
368 with HiFiasm yielded a contig assembly of 2.96 Gbp made up of 671 contigs with a contig N50 of
369 58 Mbp. The longest contig was 149 Mbp. Using the uniquely mapped Hi-C data, we were able to
370 scaffold the contigs and fix misassembled contigs (Figure 2B). Although the genome size (2.96
371 Gbp) and the largest scaffold (149 Mbp) were similar to the contig assembly, the scaffold N50 was
372 increased to 124 Mbp (N=10) after the Hi-C scaffolding. The obtained scaffold N90 was 21 Mbp
373 (N=25). The BUSCO scores of the *L. europaeus* assembly suggest it to be near-complete, with the
374 following results (%- of the total genome): Complete: 96.1% [Single copy: 93.1 %, Duplicated: 3.0
375 %], Fragmented 0.8 %, and Missing 2.9 % based on the mammalia_odb10 database.
376 To further improve the assembly, manual curation was performed, resulting in 76 scaffold
377 breaks, 81 joins and removal of 62 erroneously duplicated contigs (Table 1). These interventions
378 led to an increase in scaffold N50 of 1.69 % from 123.7 Mb to 125.8 Mb, and a reduction in scaffold
379 count of 9.21 % from 716 to 650. Of the finalized assembly, 93.16 % could be assigned to 25
380 identified chromosomes (23 autosomes plus X and Y), matching the known karyotype.
381 Chromosomes were numbered according to size (Table 2). The curated genome length was
382 2,930,972,003 bp.

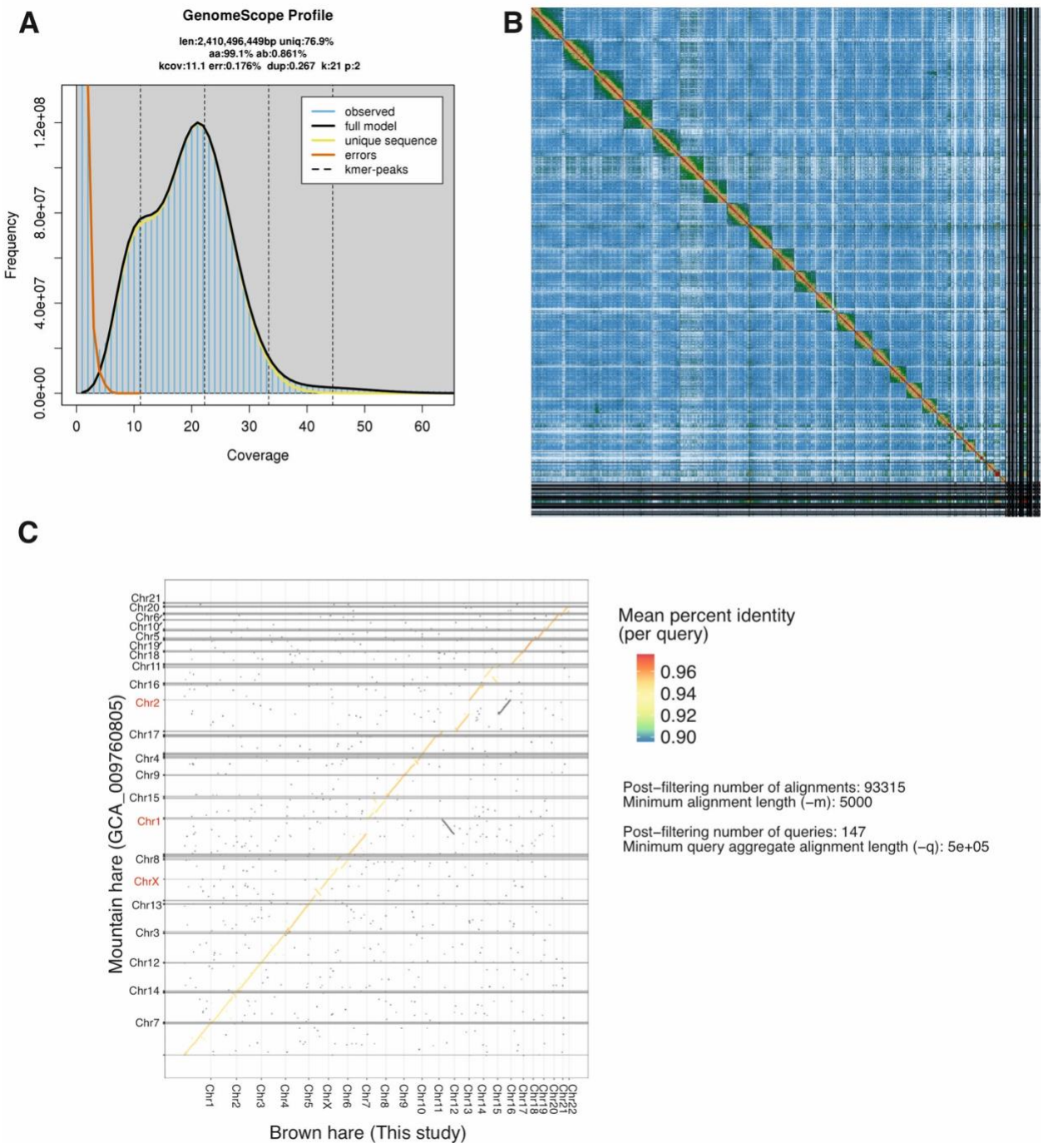

Figure 2. A) Genomescope2 profile of the PacBio HiFi data. B) YaHS pretext map of Hi-C scaffolding after manual curation. C) Dot plot comparison with the mountain hare pseudoreference (GCA_009760805) assembly with rabbit chromosomal assignments. Each grey horizontal line represents a contig end. Names of the non-chromosomal contigs have been omitted for clarity. Rabbit chromosomes corresponding two hare chromosomes as well as the X chromosome with an inversion are are highlighted.

392 **Table 1.** The brown hare reference genome chromosome assignment and assembly statistics

Chromosome assignment	
Chr length	2,730,543,680 bp
Autosomes	23
X	1
Y	1
Unlocalised	19

Statistics for primary assembly after manual curation	
# scaffolds	650
Total scaffold length	2,930,972,003
Average scaffold length	4,509,187.70
Scaffold N50	125,776,599
Scaffold auN	117,151,857.99
Scaffold L50	10
Largest scaffold	184,151,585
Smallest scaffold	1000
# contigs	1068
Total contig length	2,930,889,045
Average contig length	2,744,278.13
Contig N50	43,409,466
Contig auN	47,572,960.95
Contig L50	21
Largest contig	141,329,000
Smallest contig	1000
# gaps in scaffolds	418
Total gap length in scaffolds	82,958
Average gap length in scaffolds	198.46
Gap N50 in scaffolds	200
Gap auN in scaffolds	199.70
Gap L50 in scaffolds	208
Largest gap in scaffolds	200
Smallest gap in scaffolds	31
Base composition	
A	821,587,687
C	645,897,412
G	644,604,632
T	818,799,314

GC content	44.03 %
# soft-masked bases	0
# segments	1,068
Total segment length	2,930,889,045
Average segment length	2,744,278.13
# gaps	418
# paths	650

Table 2. Chromosome assemblies of the brown hare reference genome with their corresponding GenBank access numbers.

Chromosome	GenBank	Size (bp)	GC content (%)
1	CM065551.1	184,151,585	42.0
2	CM065552.1	176,467,698	42.0
3	CM065553.1	169,342,607	42.0
4	CM065554.1	166,606,208	43.0
5	CM065555.1	159,335,674	46.0
6	CM065556.1	132,443,831	42.0
7	CM065557.1	132,293,278	44.0
8	CM065558.1	130,587,543	41.0
9	CM065559.1	125,776,599	44.5
10	CM065560.1	125,704,418	46.0
11	CM065561.1	117,559,561	44.0
12	CM065562.1	105,667,338	44.0
13	CM065563.1	100,925,866	44.0
14	CM065564.1	99,301,477	44.0
15	CM065565.1	95,895,056	42.0
16	CM065566.1	91,637,918	42.5
17	CM065567.1	84,898,765	45.5
18	CM065568.1	77,563,730	48.5
19	CM065569.1	72,700,603	47.5
20	CM065570.1	69,331,453	45.5
21	CM065571.1	60,609,710	47.5
22	CM065572.1	44,025,767	47.0
23	CM065573.1	30,716,498	53.0
X	CM065574.1	135,183,192	40.5
Y	CM065575.1	33,560,193	41.0
MT	OR915849.1	17,540	38.6

Genome annotation

EDTA produced a curated custom repeat library of 7,045 repetitive elements. Interestingly, the proportion of the genome masked as repetitive elements was higher than expected at 46.8 % of the genome. In the mountain hare pseudoreference genome assembly (GCA_009760805), only 28 % of the genome was masked, and similarly the *k*-mer based estimate of the repetitive elements was 26.7 %. To annotate the genome, we mapped ~1.6 billion RNA-seq reads from previously published RNA-seq libraries, including this cell line. The average mapping rate of all libraries was 89.95 %, which we consider a good indicator of the quality of this genome. Using BRAKER3, we were able to annotate 19,906 gene models in the genome. We then ran BUSCO on the predicted genes and the results show a good level of completeness: Complete: 81.1 % [Single copy: 63.3 %, Duplicate: 17.8 %], Fragmented: 1.1 %, Missing: 17.8 %, Total: 9226 when compared to the mammalia_odb10 database. The final annotation (NCBI RefSeq GCF_033115175.1-RS_2024_01) contains 30,833 genes of which 21,467 code for a polypeptide. Telomeric sequences on both ends of the chromosome were detected on 13 chromosomes (Figure 3A). Furthermore, telomeric sequences were found in high copy number throughout the chromosome length, a feature previously noted using FISH in mountain hares (Forsyth et al., 2005).

Comparison to the mountain hare assembly

Minimap2 was able to align 97.3 % of the contig sequences from the mountain hare pseudoreference genome assembly (GCA_009760805) to this genome assembly (Figure 2C). As pointed out earlier, the mountain hare genome assembly has been scaffolded against the rabbit reference genome (Giska et al., 2022), which has a different chromosome count (Beklemisheva et al., 2011) and explains the few notable inversions and structural differences between the two genomes (Figure 2C). For example, the chr7 and chr 12 assemblies of the brown hare have a good degree of synteny with the mountain hare pseudoreference (= rabbit) chr1, with the chr12 being in inverted orientation. Similarly, the brown hare chr13 and chr16 assemblies have a good synteny and orientation with the rabbit chr2.

Mitochondrial genome

While it is possible to recover the entire mitochondrial genome from the HiFi sequencing data (Uliano-Silva et al., 2023), we had already previously sequenced the mitochondrial genomes of our hare cell lines for the purposes of another study. The assembly of the Illumina data provided a 16,836 bp mtDNA sequence, which was complete and circular, with a 142 bp non-repetitive overlap between the 3´ and 5´ ends of the original assembly removed from the final version. All the expected vertebrate mitochondrial genes, 37 in total, were found (Figure 3B).

Interestingly, the primary assembly of the Illumina data was notably smaller than expected based on the original PCR amplicons, with a notable difference between the expected and obtained non-coding region (NCR) length. Sanger sequencing of the PCR-fragment containing the NCR (Lt16056F + Le184R) revealed that it contained a sequence run consisting of six head-to-tail repeats of a 140 bp element between OriH and tRNA-Phe, explaining the incorrect assembly of the NCR sequence from the short-read Illumina data. With the addition of the repeat elements, the final mitochondrial genome size of our specimen is 17,540 bp.

Based on the sequence similarity with human and mouse mitochondrial genomes, we were able to tentatively identify the termination associated sequence (TAS, 2 nt difference to

human), required for the generation of the 7S DNA and thus the displacement loop (D-loop) structure on the NCR (Crews et al., 1979). Interestingly, assuming that the 7S DNA is roughly the same size as in other mammals (600-650 nt), its 5'-end would map to the beginning of another highly repetitive 240 bp region, likely to have structural significance. This is a notably longer sequence run than the tRNA-like sequence reported at the replication origin of *L. timidus* mtDNA (Melo-Ferreira et al., 2014). We have preliminary assigned this structure to correspond to the heavy-strand origin of replication (OriH), supported by the fact that a similar, tRNA sequence-derived hairpin structure is required for the priming of the origin of the light-strand replication (OriL) (Fuste et al., 2010). Experimental validation of the proposed OriH is required.

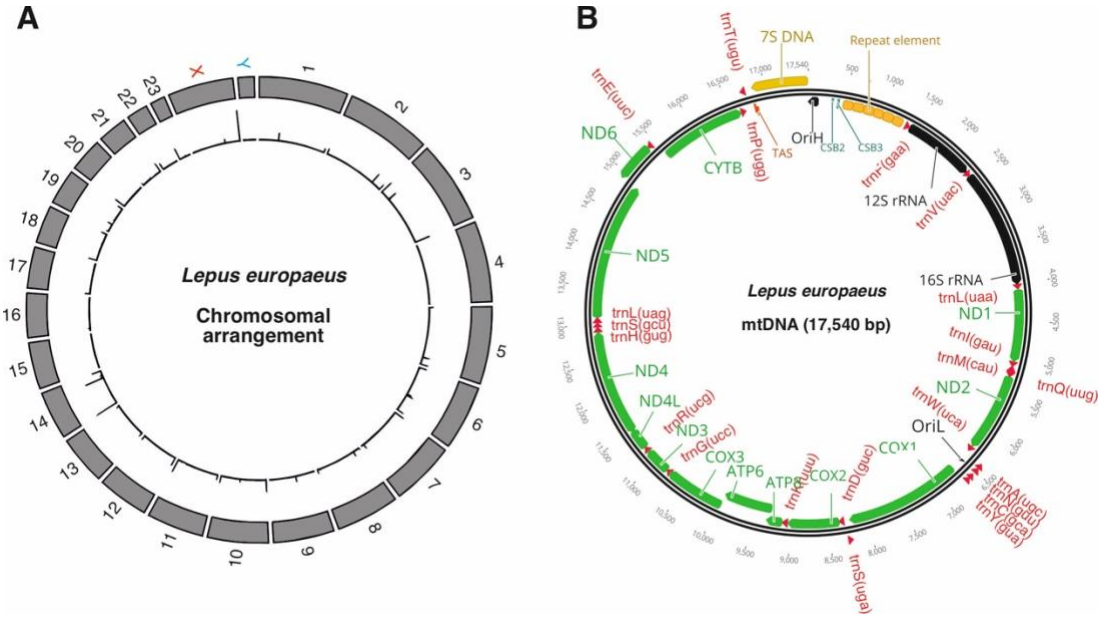

Figure 3. A reference genome for the brown hare. A) A size map of 23 autosomes as well as X and Y sex chromosomes of *Lepus europaeus*. Telomeric sequences and their relative lengths are indicated on the inner circle. B) A schematic illustration of the mountain hare mitochondrial genome. Genes encoded by the H-strand are placed inside and the L-strand genes outside of the circle. NCR = non-coding region.

Discussion

Given the challenges of assembling a large mammalian genome, we have produced a high-quality genome assembly for the brown hare. The genome assembly has a high degree of contiguity (Contig N50 43 Mbp) and completeness (BUSCO complete 96.1 %). Interestingly, the genome contains a high number of repetitive elements, a fact that warrants further investigation to elucidate their identity. While vast majority of the sequences seem to come from retroelements or endoretroviruses, we cannot fully exclude that some multicopy host genes could be masked. A detailed analysis of the repeat element libraries from this and another vertebrate genome assembly releases might be warranted.

468 The genome assembly is in chromosome scale, and all 23 autosomes as well as X and Y
469 chromosomes could be scaffolded. Compared with the closely related mountain hare
470 pseudoreference assembly, the brown hare genome assembly is much more contiguous (Table 1)
471 and inclusion of telomeric repeats on both ends of several chromosomes (Figure 3A) further
472 demonstrates the telomere-to-telomere continuity of these assemblies. The presented assembly
473 also resolves the expected chromosome structure for the species (Figure 2B, C), which is
474 informative regarding the karyotype difference between hares (n = 24) and rabbits (n = 22).
475 Comparison of the brown hare assembly with the mountain hare pseudoreference chromosomes
476 shows that the rabbit chromosome 1 corresponds to the brown hare chromosomes 12 and 17,
477 with the chromosomal arm corresponding to the former is in an inversed orientation. Similarly,
478 the rabbit chromosome 2 corresponds to the hare chromosomes 13 and 16, both in the same
479 orientation. It should also be noted that as the hare chromosomes were numbered according to
480 their size, the numbering in general does not match those of the rabbit chromosomes (Figure 2C)
481 The obtained 17,540 bp long mitochondrial genome assembly also provides some
482 interesting aspects of mtDNA sequence variation within mammalian species. The size difference
483 to human (16,569 bp) and mouse (16,298 bp) mitochondrial genomes is caused by a rather long
484 non-coding region (2,102 bp vs. 1,123 bp and 879 bp in humans and mice, respectively), longer
485 rRNA and tRNA genes, as well as additional non-coding nucleotides between genes. The
486 mitochondrial genome of our specimen is slightly smaller than the 17,734 bp previous NCBI
487 Reference Sequence for brown hare (NC_004028.1) from Sweden. The size difference is mostly
488 due to a copy number difference in tandem repeat elements within the non-coding region, with
489 seven in the NC_004028.1 and six in our specimen. Indels elsewhere in the NCR explain the rest
490 of the difference.
491 The main technical challenge for highly continuous, good quality reference genome
492 assemblies is the availability of relatively large quantities of intact HMW DNA for HiFi sequencing.
493 This typically requires snap freezing of tissue samples and assuring an intact cold chain of below
494 -80 °C for their sending, handling, and long-term storage (Januszczak and Holroyd, 2023). While
495 development in sequencing technologies can reduce the required DNA amounts, obtaining high
496 sequencing coverage over a large vertebrate genome is easier when material is plentiful. Similarly,
497 RNA is very sensitive to degradation in post-mortem tissue samples, also requiring immediate
498 preparation or snap-freezing to be suitable for RNA-seq. Although transcript data is not obligatory,
499 it is highly useful for the purpose of genome annotation. The requirement of large samples of
500 fresh tissue for -omics purposes complicates the sampling of most vertebrates species. For
501 example, in Finland all land vertebrates are protected by law (animal protection law 247/1996,
502 hunting law 615/1993), with very few exceptions, such as game animals and pest species.
503 Capturing and lethal sampling is especially problematic for endangered species. Not only will the
504 sampling require extensive permits, but it is also difficult to justify and conduct ethically. Here,
505 the generation of primary fibroblast cell lines can offer a solution. Although the source specimen
506 for the cell line utilized in this study was hunted, fibroblasts can be isolated from relatively small
507 skin biopsies, such as ear clippings (Seluanov et al., 2010), with minimally invasive sampling and
508 harm to the individual. These cells can be expanded in tissue culture and, when immortalized,
509 provide a highly scalable source of fresh DNA and RNA. Furthermore, living cells can be stored as
510 frozen stocks in biobanks for decades (Fazekas et al., 2017). Tissue sampling can be done directly
511 into the growth medium in ambient temperature, allowing several days for the transport to the

laboratory without a need for dry ice or N_2-cryocontainers, which greatly simplifies collection and logistics. The same applies to the isolated cells. As an example, we sent living cell suspensions from Joensuu, Finland to the ZFMK biobank in Bonn, Germany, over regular airmail in ambient temperature. The receiving laboratory can then amplify the cells by culturing and cryopreserve them in sufficient replicate stocks. It is worth to point out that the obtained genome assembly scaffolded to 23 autosomes, X and Y chromosomes, as expected for the species, shows no evidence of karyotypic drift caused by cell culture, dispelling possible concerns for the use of cell lines for such work (Gaertner et al., 2022).

The brown hare is an iconic European mammal, locally familiar to a wider public of nature-goers, farmers and hunters, while being at the same time an invasive species on many other continents. It has populations with complex ancestral makeup due to isolated Pleistocene refugia, frequent introductions and cross-species hybridization, making it an exceptionally interesting mammal for genetic studies. Furthermore, depending on the local environmental context, brown hares can be both an endangered as well as a rapidly expanding species, prompting opportunities for both conservation as well as invasion genetics studies. A high-quality reference genome assembly will allow us also to peer deeper into genome structure and chromosome evolution, which will be particularly interesting from the viewpoint of genetic introgression (Juric et al., 2016) and maintenance of species boundaries. Any issues of genetic compatibility between species should leave their fingerprint in the genomes (Skov et al., 2023), providing interesting insight into the speciation mechanisms (Wolf et al., 2010). Having a high-quality genome for both species has important implications for the fields of evolutionary and conservation biology. Due to the high contiguity of the genome assembly, it will be possible to perform accurate analyses of linkage disequilibrium as well as the identification of runs of homozygosity (ROH). Accurate ROH analysis is useful for identifying and understanding the impact of inbreeding burden on the genome of increasingly threatened and isolated species, as quantitative traits can be associated with ROH burden (Ceballos et al., 2018). Furthermore, this information might provide important insights into the hybridization of the brown hare with the mountain hare, as well as help to pinpoint genomic regions that might be helping the species to adapt to local environmental conditions during the ongoing range expansion (Pohjoismäki et al., 2021). Other *Lepus* reference genome assemblies available soon will help to complete this picture. Together, this genome assembly and the mountain hare genome assembly will provide a solid base for future studies.

Acknowledgements

We acknowledge the DNA Sequencing and Genomics Laboratory, Institute of Biotechnology, University of Helsinki for the PacBio and Hi-C sequencing, with special thanks to Dr Petri Auvinen and Dr Martyn James for helpful discussions and personal interest towards the project. 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 are grateful to the European Reference Genome Atlas (ERGA) for the opportunity to include our mountain hare genome into their pilot project. Dr Ann McCartney is thanked for her kind help throughout the process and Mr Jukka Pusa for his contribution in obtaining the brown hare specimen. We thank Ms Anita

Kervinen for her valuable laboratory assistance and Dr. Mervi Kunnasranta for kindly allowing to use her hare photo in the publication.

Data, scripts, code, and supplementary information availability

The genome assembly is available from the NCBI database under BioProject ID PRJNA950335. The cell line is available through the corresponding author as well as from the ZFMK biobank.

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

Funding

This study belongs to the xHARES consortium funded by the R'Life initiative of the Academy of Finland, grant no. 329264.

References

10x-Genomics, 2017. Salting out method for DNA extraction from cells. Sample Preparation Demostrated Protocol. CG00016. Rev A., 1-5.

Bai, Y., Lin, W., Xu, J., Song, J., Yang, D., Chen, Y.E., Li, L., Li, Y., Wang, Z., Zhang, J., 2021. Improving the genome assembly of rabbits with long-read sequencing. Genomics 113, 3216-3223.

Beklemisheva, V.R., Romanenko, S.A., Biltueva, L.S., Trifonov, V.A., Vorobieva, N.V., Serdukova, N.A., Rubtsova, N.V., Brandler, O.V., O'Brien, P.C., Yang, F., Stanyon, R., Ferguson-Smith, M.A., Graphodatsky, A.S., 2011. Reconstruction of karyotype evolution in core Glires. I. The genome homology revealed by comparative chromosome painting. Chromosome Res 19, 549-565.

Belaghzal, H., Dekker, J., Gibcus, J.H., 2017. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. Methods 123, 56-65.

Birney, E., Clamp, M., Durbin, R., 2004. GeneWise and Genomewise. Genome Res 14, 988-995.

Blaxter, M., Archibald, J.M., Childers, A.K., Coddington, J.A., Crandall, K.A., Di Palma, F., Durbin, R., Edwards, S.V., Graves, J.A.M., Hackett, K.J., Hall, N., Jarvis, E.D., Johnson, R.N., Karlsson, E.K., Kress, W.J., Kuraku, S., Lawniczak, M.K.N., Lindblad-Toh, K., Lopez, J.V., Moran, N.A., Robinson, G.E., Ryder, O.A., Shapiro, B., Soltis, P.S., Warnow, T., Zhang, G.J., Lewin, H.A., 2022. Why sequence all eukaryotes? P Natl Acad Sci USA 119.

Bock, A., 2020. Lepus europaeus (Lagomorpha: Leporidae). Mammalian Species 52, 125-142.

591 Bruna, T., Hoff, K.J., Lomsadze, A., Stanke, M., Borodovsky, M., 2021. BRAKER2: automatic
592 eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein
593 database. NAR Genom Bioinform 3, lqaa108.
594 Carneiro, M., Rubin, C.J., Di Palma, F., Albert, F.W., Alfoldi, J., Martinez Barrio, A., Pielberg, G.,
595 Rafati, N., Sayyab, S., Turner-Maier, J., Younis, S., Afonso, S., Aken, B., Alves, J.M., Barrell, D.,
596 Bolet, G., Boucher, S., Burbano, H.A., Campos, R., Chang, J.L., Duranthon, V., Fontanesi, L.,
597 Garreau, H., Heiman, D., Johnson, J., Mage, R.G., Peng, Z., Queney, G., Rogel-Gaillard, C.,
598 Ruffier, M., Searle, S., Villafuerte, R., Xiong, A., Young, S., Forsberg-Nilsson, K., Good, J.M.,
599 Lander, E.S., Ferrand, N., Lindblad-Toh, K., Andersson, L., 2014. Rabbit genome analysis
600 reveals a polygenic basis for phenotypic change during domestication. Science 345, 1074-
601 1079.
602 Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M., Wilson, J.F., 2018. Runs of homozygosity:
603 windows into population history and trait architecture. Nat Rev Genet 19, 220-234.
604 Chen, S.F., Zhou, Y.Q., Chen, Y.R., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor.
605 Bioinformatics 34, 884-890.
606 Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., Li, H., 2021. Haplotype-resolved de novo
607 assembly using phased assembly graphs with hifiasm. Nat Methods 18, 170-175.
608 Crews, S., Ojala, D., Posakony, J., Nishiguchi, J., Attardi, G., 1979. Nucleotide sequence of a region
609 of human mitochondrial DNA containing the precisely identified origin of replication.
610 Nature 277, 192-198.
611 Fazekas, J., Grunt, T.W., Jensen-Jarolim, E., Singer, J., 2017. Long term storage in liquid nitrogen
612 leads to only minor phenotypic and gene expression changes in the mammary carcinoma
613 model cell line BT474. Oncotarget 8, 35076-35087.
614 Ferreira, M.S., Jones, M.R., Callahan, C.M., Farelo, L., Tolesa, Z., Suchentrunk, F., Boursot, P., Mills,
615 L.S., Alves, P.C., Good, J.M., Melo-Ferreira, J., 2021. The Legacy of Recurrent Introgression
616 during the Radiation of Hares. Syst Biol 70, 593-607.
617 Fickel, J., Hauffe, H.C., Pecchioli, E., Soriguer, R., Vapa, L., Pitra, C., 2008. Cladogenesis of the
618 European brown hare (Lepus europaeus Pallas, 1778). Eur J Wildlife Res 54, 495-510.
619 Forsyth, N.R., Elder, F.F., Shay, J.W., Wright, W.E., 2005. Lagomorphs (rabbits, pikas and hares) do
620 not use telomere-directed replicative aging in vitro. Mech Ageing Dev 126, 685-691.
621 Fuste, J.M., Wanrooij, S., Jemt, E., Granycome, C.E., Cluett, T.J., Shi, Y., Atanassova, N., Holt, I.J.,
622 Gustafsson, C.M., Falkenberg, M., 2010. Mitochondrial RNA polymerase is needed for
623 activation of the origin of light-strand DNA replication. Mol Cell 37, 67-78.
624 Gaertner, K., Michell, C., Tapanainen, R., Goffart, S., Saari, S., Soininmaki, M., Dufour, E.,
625 Pohjoismaki, J.L.O., 2022. Molecular phenotyping uncovers differences in basic
626 housekeeping functions among closely related species of hares (Lepus spp., Lagomorpha:
627 Leporidae). Mol Ecol.
628 Gertz, E.M., Yu, Y.K., Agarwala, R., Schaffer, A.A., Altschul, S.F., 2006. Composition-based statistics
629 and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol 4,
630 41.
631 Ghurye, J., Rhie, A., Walenz, B.P., Schmitt, A., Selvaraj, S., Pop, M., Phillippy, A.M., Koren, S., 2019.
632 Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput
633 Biol 15, e1007273.

634 Giska, I., Pimenta, J., Farelo, L., Boursot, P., Hacklander, K., Jenny, H., Reid, N., Montgomery, W.I.,
635 Prodohl, P.A., Alves, P.C., Melo-Ferreira, J., 2022. The evolutionary pathways for local
636 adaptation in mountain hares. Mol Ecol 31, 1487-1503.
637 Gustavsson, I., 1972. Mitotic and meiotic chromosomes of the variable hare (Lepus timidus L.),
638 the common hare (Lepus europaeus Pall.) and their hybrids. Hereditas 67, 27-34.
639 Holden, M.E., 2005. Family Gliridae, in: Wilson, D.E., Reeder, D.M. (Eds.), Mammal Species of the
640 World: A Taxonomic and Geographic Reference. Johns Hopkins University Press, Baltimore,
641 USA, pp. 819-841.
642 Howe, K., Chow, W., Collins, J., Pelan, S., Pointon, D.-L., Sims, Y., Torrance, J., Tracey, A., Wood, J.,
643 2021. Significantly improving the quality of genome assemblies through curation.
644 GigaScience 10.
645 Jaksic, F.M., 2023. Historical ecology and current status of the European hare Lepus europaeus in
646 South America: a new species in new countries. Stud Neotrop Fauna E.
647 Januszczak, I., Holroyd, N., 2023. DToL Tissue and Blood Sampling Standard Operating Procedure:
648 Chordata: Vertebrata (general). protocols.io
649 dx.doi.org/10.17504/protocols.io.n92ldmjool5b/v1.
650 Juhling, F., Putz, J., Bernt, M., Donath, A., Middendorf, M., Florentz, C., Stadler, P.F., 2012.
651 Improved systematic tRNA gene annotation allows new insights into the evolution of
652 mitochondrial tRNA structures and into the mechanisms of mitochondrial genome
653 rearrangements. Nucleic Acids Res 40, 2833-2845.
654 Juric, I., Aeschbacher, S., Coop, G., 2016. The Strength of Selection against Neanderthal
655 Introgression. PLoS Genet 12, e1006340.
656 Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment
657 and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37, 907-915.
658 Korstanje, R., O'Brien, P.C., Yang, F., Rens, W., Bosma, A.A., van Lith, H.A., van Zutphen, L.F.,
659 Ferguson-Smith, M.A., 1999. Complete homology maps of the rabbit (Oryctolagus
660 cuniculus) and human by reciprocal chromosome painting. Cytogenet Cell Genet 86, 317-
661 322.
662 Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A.,
663 2009. Circos: an information aesthetic for comparative genomics. Genome Res 19, 1639-
664 1645.
665 Lawniczak, M.K.N., Durbin, R., Flicek, P., Lindblad-Toh, K., Wei, X., Archibald, J.M., Baker, W.J.,
666 Belov, K., Blaxter, M.L., Marques Bonet, T., Childers, A.K., Coddington, J.A., Crandall, K.A.,
667 Crawford, A.J., Davey, R.P., Di Palma, F., Fang, Q., Haerty, W., Hall, N., Hoff, K.J., Howe, K.,
668 Jarvis, E.D., Johnson, W.E., Johnson, R.N., Kersey, P.J., Liu, X., Lopez, J.V., Myers, E.W.,
669 Pettersson, O.V., Phillippy, A.M., Poelchau, M.F., Pruitt, K.D., Rhie, A., Castilla-Rubio, J.C.,
670 Sahu, S.K., Salmon, N.A., Soltis, P.S., Swarbreck, D., Thibaud-Nissen, F., Wang, S., Wegrzyn,
671 J.L., Zhang, G., Zhang, H., Lewin, H.A., Richards, S., 2022. Standards recommendations for
672 the Earth BioGenome Project. Proc Natl Acad Sci U S A 119.
673 Levänen, R., Kunnasranta, M., Pohjoismäki, J., 2018a. Mitochondrial DNA introgression at the
674 northern edge of the brown hare (Lepus europaeus) range. Ann Zool Fenn 55, 15-24.
675 Levänen, R., Pohjoismäki, J., Kunnasranta, M., 2019. Home ranges of semi-urban brown hares
676 (Lepus europaeus) and mountain hares (Lepus timidus) at northern latitudes. Ann Zool Fenn
677 56, 107–120.

Levänen, R., Thulin, C.G., Spong, G., Pohjoismäki, J.L.O., 2018b. Widespread introgression of mountain hare genes into Fennoscandian brown hare populations. Plos One 13.

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

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094-3100.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., Dekker, J., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289-293.

Manni, M., Berkeley, M.R., Seppey, M., Zdobnov, E.M., 2021. BUSCO: Assessing Genomic Data Quality and Beyond. Curr Protoc 1, e323.

Marques, J.P., Seixas, F.A., Farelo, L., Callahan, C.M., Good, J.M., Montgomery, W.I., Reid, N., Alves, P.C., Boursot, P., Melo-Ferreira, J., 2020. An Annotated Draft Genome of the Mountain Hare (Lepus timidus). Genome Biol Evol 12, 3656-3662.

Mazzoni, C.J., Ciofi, C., Waterhouse, R.M., 2023. Biodiversity: an atlas of European reference genomes. Nature 619, 252.

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

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

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., Gurevich, A., 2018. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 34, i142-i150.

Minoudi, S., Papapetridis, I., Karaiskou, N., Chatzinikos, E., Triantaphyllidis, C., Abatzopoulos, T.J., Triantafyllidis, A., 2018. Genetic analyses of brown hare (Lepus europaeus) support limited migration and translocation of Greek populations. Plos One 13, e0206327.

Ognev, S.I., 1940. 157. Lepus (Eulagus) europeaus Pall. (1778) заяц-русак, Mammals of the USSR and adjacent countries: rodents (In Russian). Akademiya Nauk SSSR, Moscow, pp. 112-161.

Open2C, Abdennur, N., Fudenberg, G., Flyamer, I.M., Galitsyna, A.A., Goloborodko, A., Imakaev, M., Venev, S.V., 2023. Pairtools: from sequencing data to chromosome contacts. bioRxiv.

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., Jiang, N., Hirsch, C.N., Hufford, M.B., 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol 20, 275.

Petrovan, S., 2013. Lepus europaeus (European hare), CABI Compendium.

Pohjoismäki, J.L.O., Michell, C., Levänen, R., Smith, S., 2021. Hybridization with mountain hares increases the functional allelic repertoire in brown hares. Sci Rep 11, 15771.

Ranallo-Benavidez, T.R., Jaron, K.S., Schatz, M.C., 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun 11, 1432.

722 Reid, N., 2011. European hare (Lepus europaeus) invasion ecology: implication for the
723 conservation of the endemic Irish hare (Lepus timidus hibernicus). Biol Invasions 13, 559-
724 569.
725 Reid, N., Montgomery, W.I., 2007. Is naturalisation of the brown hare in Ireland a threat to the
726 endemic Irish hare? Biol Environ 107b, 129-138.
727 Rikala, K., 1925. Ajojahdilla [flush hunting of hares]. Metsästys ja Kalastus 2, 44-48.
728 Salonius, A.H., 1924. Rusakkojäniksen leviämisalue Suomessa [Distribution of the brown hare in
729 Finland]. In Finnish. Metsästys ja Kalastus, 165-169.
730 Schai-Braun, S.C., Schwienbacher, S., Smith, S., Hackländer, K., 2023. Coexistence of European
731 hares and Alpine mountain hares in the Alps: what drives the occurrence and frequency of
732 their hybrids? Journal of Zoology 320, 214-225.
733 Schai-Braun, S.C., Weber, D., Hacklander, K., 2013. Spring and autumn habitat preferences of
734 active European hares (Lepus europaeus) in an agricultural area with low hare density. Eur
735 J Wildlife Res 59, 387-397.
736 Seluanov, A., Vaidya, A., Gorbunova, V., 2010. Establishing primary adult fibroblast cultures from
737 rodents. J Vis Exp.
738 Siivonen, L., 1972. Rusakko. Kustannusosakeyhtiö Otava, Keuruu, Finland.
739 Skov, L., Coll Macia, M., Lucotte, E.A., Cavassim, M.I.A., Castellano, D., Schierup, M.H., Munch, K.,
740 2023. Extraordinary selection on the human X chromosome associated with archaic
741 admixture. Cell Genom 3, 100274.
742 Smit, A.F.A., Hubley, R., Green, P., 2013-2015. RepeatMasker Open-4.0.
743 http://www.repeatmasker.org.
744 Smith, R.K., Jennings, N.V., Harris, S., 2005. A quantitative analysis of the abundance and
745 demography of European hares Lepus europaeus in relation to habitat type, intensity of
746 agriculture and climate. Mammal Rev 35, 1-24.
747 Stott, P., 2003. Use of space by sympatric European hares (Lepus europaeus) and European rabbits
748 (Oryctolagus cuniculus) in Australia. Mamm Biol 68, 317-327.
749 Suomalainen, E.W., 1922. Rusakkojäniksen leveneminen Suomessa [Distribution of the brown
750 hare in Finland]. In Finnish. Metsästys ja Kalastus 6, 165-169.
751 Thulin, C.-G., 2003. The distribution of mountain hares (Lepus timidus, L. 1758) in Europe: A
752 challenge from brown hares (L. europaeus, Pall. 1778)? Mammal Rev 33, 29-42.
753 Thulin, C.-G., Tegelström, H., 2002. Biased geographical distribution of mitochondrial DNA that
754 passed the species barrier from mountain hares to brown hares (genus Lepus): an effect of
755 genetic incompatibility and mating behaviour? J Zool 258, 299-306.
756 Thulin, C.G., Winiger, A., Tallian, A.G., Kindberg, J., 2021. Hunting harvest data in Sweden indicate
757 precipitous decline in the native mountain hare subspecies Lepus timidus sylvaticus (heath
758 hare). J Nat Conserv 64.
759 Uliano-Silva, M., Ferreira, J., Krasheninnikova, K., Darwin Tree of Life, C., Formenti, G., Abueg, L.,
760 Torrance, J., Myers, E.W., Durbin, R., Blaxter, M., McCarthy, S.A., 2023. MitoHiFi: a python
761 pipeline for mitochondrial genome assembly from PacBio high fidelity reads. BMC
762 Bioinformatics 24, 288.
763 Vinogradov, A.E., 1998. Genome size and GC-percent in vertebrates as determined by flow
764 cytometry: The triangular relationship. Cytometry 31, 100-109.

765 Wadén, D.J., 1910. Rusakkojäniksen (Lepus europaeus) nykyisestä leviämisestä Suomessa [The
766 current distribution of the brown hare in Finland]. In Finnish. Suomen metsästyslehti 8, 204-
767 207.
768 Wheeler, T.J., Eddy, S.R., 2013. nhmmer: DNA homology search with profile HMMs. Bioinformatics
769 29, 2487-2489.
770 Wolf, J.B., Lindell, J., Backstrom, N., 2010. Speciation genetics: current status and evolving
771 approaches. Philos Trans R Soc Lond B Biol Sci 365, 1717-1733.
772 Zhou, C., McCarthy, S.A., Durbin, R., 2023. YaHS: yet another Hi-C scaffolding tool. Bioinformatics
773 39.
774