THESIS FOR DOCTORAL DEGREE (PH.D.)

ZSÓFIA FEKETE

GÖDÖLLŐ

2025

Magyar Agrár- és Élettudományi Egyetem

GENOMIC INVESTIGATIONS OF LEPORIDAE SPECIES AND MITOCHONDRIAL SEQUENCES THROUGH BIOINFORMATIC METHODS

Zsófia Fekete

2025.

Gödöllő

The Doctoral School:

Name: **Állatbiotechnológiai** és **Állattudományi Doktori Iskola**

Discipline: Animal Biotechnology

Head of Doctoral School: **Prof. Dr. Mézes Miklós (D.Sc)**

Professor, member of HAS

Hungarian University of Agriculture and Life Sciences

Institute of Physiology and Nutrition, Department of Feed Safety

Supervisors: **Dr. Kontra Levente**

PhD, Leader of the Bioinformatics Core Facility of

HUN-REN; Institute of Experimental Medicine (KOKI)

Bioinformatics Central Unit

Dr. Hoffmann Orsolya Ivett

PhD, senior research fellow, group leader

Hungarian University of Agriculture and Life Sciences

Institute of Genetics and Biotechnology, Department of Animal Biotechnology

.....

Head of the Doctoral School

Prof. Dr. Miklós Mézes

.....	.....
Supervisor	Supervisor
Dr. Levente Kontra	Dr. Orsolya Ivett Hoffmann

Table of Contents

1. Introduction and aim of the work

Leporidae species can be found all over the world, populating various geographical locations and fulfilling diverse ecological roles. The European rabbit (*Oryctolagus cuniculus*) is the only currently extant species of the genus *Oryctolagus*, however, it exists in both its domesticated and wild forms. On the other hand, the genus *Lepus* consists of numerous extant species worldwide. Despite the worldwide presence of members of the genus and the fact that they have occasionally been associated with humans (Somerville *et al.*, 2017; Sheng *et al.*, 2020), they were never domesticated.

Population genomics, and the research of selection is an important topic for animal husbandry as well as for evolutionary and ecological research and conservation. Using modern sequencing technologies, we can examine selection, hybridization and introgression with high resolution and accuracy. Natural and artificial selection leaves behind similar genomic signatures, thus, they can be analyzed through similar methods. Reference genomes – and their quality – are playing an important role in detecting these. The reference genome used has an influence on the quality of the read alignment and subsequently on the variants called and the end results of any analysis (Nevado *et al.*, 2014; Bohling, 2020). So-called draft genomes have been generated for numerous non-model species; however, these are now superseded by highly continuous, chromosome-level assemblies, which are important in minimizing biases caused by the quality of the reference.

Mitochondrial markers are widely used in phylogenetic and population genetic experiments, although mitochondrial introgression is fairly common in hybridizing populations. This introgression may be advantageous for the receiving population, but it is also a confounding factor in evolutionary research. The potential recombination of mtDNA is known in a number of species and taxa

as well, even on an evolutionary timescale (Ladoukakis and Zouros, 2001). The mtDNA of mammals is maternally inherited, therefore even if recombination happens, it would not have detectable effect in most cases. At the same time, heteroplasmic cell lines and individuals exist, and in some cases, it is quite prevalent, even between hybridizing species, such as the *Lepus timidus* and *Lepus europaeus* (Tapanainen *et al.*, 2024). This way, we can't exclude the possibility of evolutionary importance of mtDNA recombination.

1.1. The aims of this work

1.1.1. Population genomic analyses in rabbit (Oryctolagus cuniculus)

Despite the large amount of knowledge about rabbits, we know little of the origin and current state of wild rabbit populations of Central Europe, including those of Hungary. My goal was to examine the current situation of wild rabbit populations in Hungary, examining the possibility of admixture with domestic rabbits and if there were signs of inbreeding.

Next to this, I examined selection during domestication. I was looking for selected regions and genetic variants, the amount and genomic positions of these, while also looking at their possible roles.

I used publicly available whole genome sequencing data as well as data from newly sequenced samples. Besides the samples originating in Hungary, and sequenced by us for this project, I included samples available in the NCBI SRA database, originating from the Iberian Peninsula and France to achieve a better picture of the state of the examined populations.

1.1.2. Assembling the reference genome of the mountain hare (Lepus timidus)

I aimed to create a new reference genome for the mountain hare (*Lepus timidus*) that is up to modern standards, and thus can function as a reliable reference genome for future research. For this, I followed the current best practices, established by international genome assembly consortia (Rhie *et al*.,

2020a). Following an assembly, I compared the newly assembled genome to a closely related species with which the mountain hare is able to hybridize, the brown hare (*Lepus europaeus*)

1.1.3. Creating a bioinformatic method for identifying recombinant mitochondrial sequences

My third goal was to develop a bioinformatics method to detect recombination of mitochondrial DNA sequences at a read level. For this, we generated highly accurate long read sequencing data (PacBio HiFi, Wenger *et al*., 2019).

Even though there are numerous extant methods of identifying recombination, these usually focus on meiotic recombination, or on prokariotic and viral genomes where the prevalence of recombination is known and accepted. Despite evidence of it in some heteroplasmic cases (Kraytsberg *et al*., 2004; Zsurka *et al*., 2005, 2007), recombination of mammalian mtDNA has not yet been widely accepted.

My method's main goal is to support molecular laboratory results, and to help prove mtDNA recombination may happen in mammalian cells.

2. Materials and methods

2.1. Whole genome sequencing based population genetic analysis of wild and domestic rabbits

We have sequenced wild and various domestic rabbit breeds as well as collected publicly available sequencing data, 100 samples in total. Our wild rabbit samples are from Hungary, France, and the Iberian Peninsula, the first group sequenced by us, while the latter two were acquired from the NCBI SRA public database. Besides these, we included various domestic breeds of both public data and those sampled and sequenced by us. After quality control and filtering using the best practices outlined for the GATK ToolKit (Auwera and O'Connor, 2020),

we produced a high-quality variant set in which each point we used for further analysis was genotyped for each sample in the set.

We used the SNPhylo (Lee *et al.*, 2014) pipeline to construct a phylogenetic tree and plink 1.9 (Chang *et al.*, 2015) for principal component analysis. Following this, we examined the possibility of admixture between domestic and wild European rabbit population, using the ADMIXTURE (Alexander *et al.*, 2009) and Treemix (Pickrell and Pritchard, 2012) programs. These results were supported by D and f4 tests as well. Runs of homozygosity were examined for each individually sequenced genome to assess the degree of inbreeding, using plink 1.9.

To examine the effects of domestication and putatively selected genomic regions, we excluded any potentially admixed samples. Signatures of selection were examined using three methods, nucleotide diversity (π), fixation index (Fst), and CLR (composite likelihood ratio). For the former two, we used vcftools (Danecek *et al.*, 2011), for the latter, the SweepFinder2 (DeGiorgio *et al.*, 2016) program. We considered the extrememost values (99^{th} percentile) of the results as potentially selected. We further examined the regions where the extremes of the 3 tests overlap. Genes overlapping with these regions were identified using the Ensembl annotation (release 104) (K. L. Howe *et al.*, 2021), while transcription binding sites were identified using LiftOver (Hinrichs, 2006) and the ChipSummitDB (Czipa *et al.*, 2020) database.

2.2. New, high-quality reference genome for the mountain hare (Lepus timidus)

The mountain hare specimen used for the assembly of the reference genome was a young male individual, hunted in Finland during hunting season. Species identity and non-hybrid status was confirmed based on both morphological and genetical markers (Fekete *et al*., 2025a; Gaertner *et al*., 2023). Nuclear DNA for the PacBio HiFi and Illumina Hi-C genome sequencing was isolated from an immortalized fibroblast cell line of the individual. For the

mitochondrial genome assembly, DNA was isolated from ear tissue of the same individual, then PCR amplified and sequenced using Illumina method.

During the genome assembly we followed protocols up to current standard (Rhie *et al.*, 2020a). Following quality control, contigging was done with Hifiasm (Cheng *et al.*, 2021), sequences were deduplicated with purde_dups (Guan *et al.*, 2020), scaffolded with YaHS (Zhou *et al.*, 2023), and finally, manually curated. Repetitive sequences were annotated and mased using RepeatModeler (Flynn *et al.*, 2020) and RepeatMasker (Smit *et al.*, 2015), while telomer sequences were identified with tidk (Brown *et al.*, 2025). The mitochondrial genome was assembled with the MitoZ (Meng *et al.*, 2019) pipeline. For quality control and assessing the completeness of the genome, we used BUSCO (Manni *et al.*, 2021) – for gene content, presence of expected orthologs – gfastats (Formenti *et al.*, 2022) and QUAST (Gurevich *et al.*, 2013) – for sequence properties and statistics.

2.3. Detecting the recombination of mitochondrial DNA in mammalian cell cultures

We used two heteroplasmic cell lines, each with two distinct haplotypes, to detect potential recombination of mtDNA. Identification of the haplogroups was done by a PhyloTree Build based online tool (Van Oven and Kayser, 2009). The cell lines used were a mouse (HP202B) and a human (BH10) heteroplasmic cell line. We used MGME1 knockout cell lines, to lessen the effectiveness of linear mtDNA degradation (Peeva *et al.*, 2018) and thus pushing the equilibrium towards recombination. Mitochondrial DNA was isolated and then sequenced using PacBio HiFi technology.

The PacBio HiFi sequencing is an amplification-free method, where one molecule is represented by one read. The high accuracy and the known haplotypes made it possible to assign haplotypes to each read based on consistently present variations.

Purely mitochondrial reads were identified by making use of the length of reads. Reads aligning completely and only to mitochondrial sequences, while also approximately the length of a mitochondrial genome, can be reliably considered mitochondrial sequences.

To identify variants present in haplotypes, I treaded mtDNA reads as if coming from a diploid genome, and used variants called as heterozygous. Sequences were aligned to one of the two haplotypes present in the cell lines, and variant calling was done using the GATK ToolKit's (Auwera and O'Connor, 2020) VariantCaller and pbsv (https://github.com/PacificBiosciences/pbsv).

Using the variants belonging to either haplotype (marking them as A or B) I generated a pseudo-sequence and set up an adjustable scoring system to assess the results. Each step was conducted in parallel for both cell lines, along with visual control steps. Scripts written and used for these projects are available at https://zenodo.org/records/10473729.

3.Results and discussion

3.1. Hybridization and selection between the wild and domestic European rabbit

The Hungarian populations of the European rabbit are genetically distinct from the Iberian and French population. Their position on the phylogenetic tree is neighbouring those and placed farther away from the domestic breeds (**Figure 1.**), suggesting their natural origins rather than being descendants of released or escaped domestic animals.

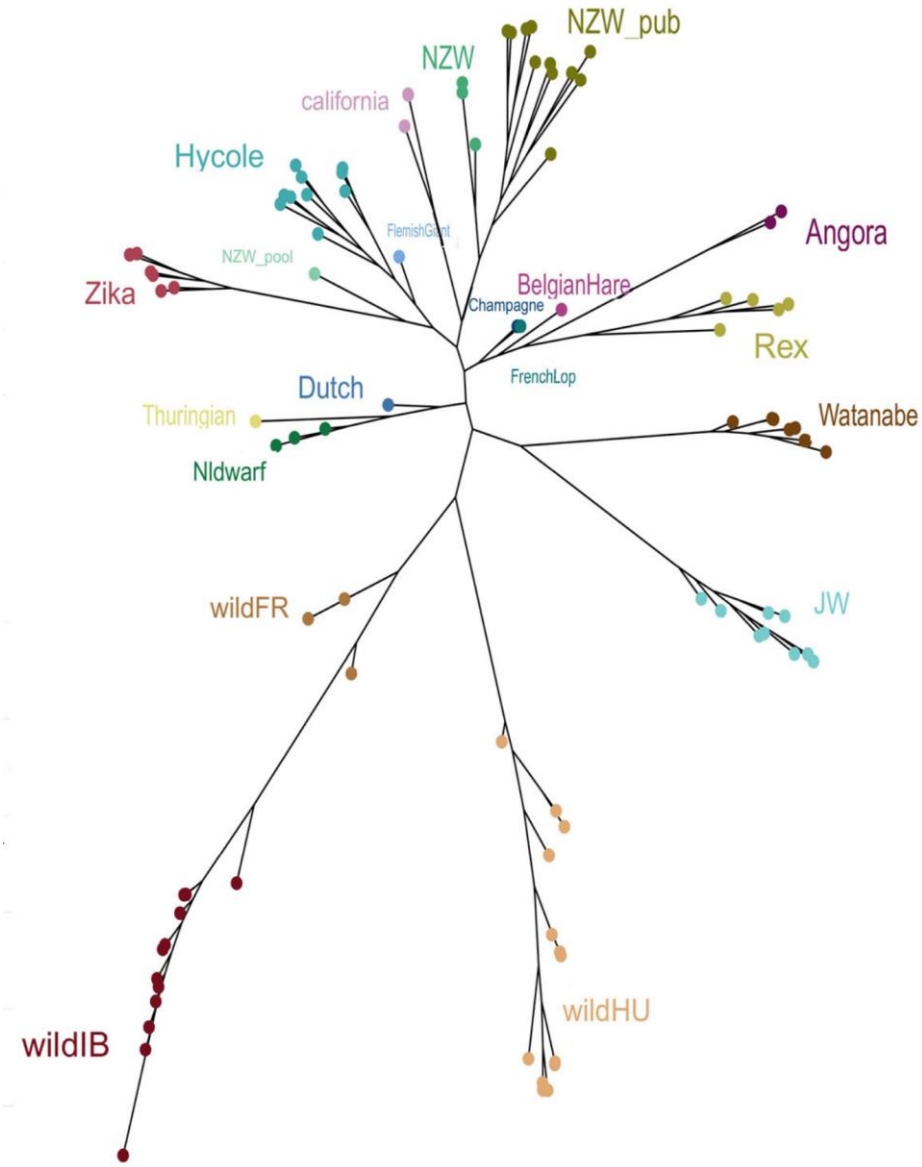

Figure 1 *– Phylogenetic tree constructed with a maximum-likelihood method. Domestic rabbit breeds are mostly grouping according to the purpose of the breed, while wild rabbits are showing geographical isolation. The two Japanese breeds are notably separate from the rest of the domestic rabbits. Abbreviations: NZW – New Zealand White; Nldward – Netherlands dwarf; JW – Japanese White; wildFR – Wild rabbit from France; wildHU – Wild rabbit from Hungary; wildIB – Wild rabbit from the Iberian Peninsula. (Fekete et al., 2025b)*

We observed the admixture of domestic and wild populations in a few cases. The admixed samples of sequenced individuals were all part of the population, currently living in the Budapest Zoo (**Figure 2.**), but we have no further information on them, therefore, we cannot conclude the time and location of admixture. Overall, we know little of the populations of the European rabbit in Hungary, as most research done in wild rabbits focuses either on their home territories of the Iberian Peninsula, or places where they are present as an invasive species. On the other hand, research aimed at domestic rabbits is mainly focused on selection between breeds.

Domestication is polygenic in its origins, based on shifts in allele frequency, rather than the emergence of new mutations. While assessing differentiation and selection between domestic and wild rabbits with three methods, we found 46 putatively selected regions. The genes overlapping with these included genes generally associated with traits that often undergo changes during domestication, such as development and growth, neural signal transduction, behaviour, and immune response.

We found a large amount of strongly differentiated (Fst>0.75) variants (983 in total) in the selected regions, but only few fixed at one allele in either domestic or wild populations. We also detected some regions (14 in total) as selected, that did not overlap with known genes. Examining the presence of transcription factor binding sites (TFBS) we found that certain types were enriched in the selected regions (**Table 1.**), showing the importance of regulatory regions.

Figure 2 *– ADMIXTURE results. K=5 and K=7 ancestral populations gave the best cross-validation error values. The two groupings resulted in similar groupings, the main difference being the New-Zealand White (NZW and NZW-public) and Zika samples forming a separate group from the rest of the domestic breeds. In the case of K=6 ancestral populations, the Iberian and French samples showed more structure. In all cases, there is a similar level of admixture detected between domestic and wild rabbits of Hungary and France. (Fekete et al., 2025b)*

Table 1 *– Transcription factor binding site (TFBS) types enriched in selected regions. First column is the total amount of the TFBS type identified on the OryCun2.0. rabbit reference genome, while the third column shows the amount in selected regions.*

All	Name	In selected region	Fisher exact
9610	**Atf3**	11	2.972E-02
1796	**Bcl6**	4	2.263E-02
46581	**CEBPB**	38	3.049E-02
5	**ESR1**	1	2.941E-03
18173	**NFIC**	21	3.292E-03
9848	**Rxra**	11	3.446E-02
2302	**ZNF143**	7	5.070E-04

3.2.A high-quality reference genome of the mountain hare (Lepus timidus)

We produced a highly continuous, chromosome-level genome assembly of the mountain hare (*Lepus timidus*), up to current standards. While the 2,7 Gbp genome is about 500 Mbp shorter than the flow-cytometry based first estimation by Vinogradov (1998), it is very close to the draft genome of the mountain hare (Marques *et al.*, 2020). The genome contains a large amount of repetitive sequences (42,35%) (Fekete, *et al.*, 2025a), similarly to the brown hare (*Lepus europaeus*) (46%) (Michell *et al.*, 2024b). This is much higher than the earlier estimation (23%) (Marques *et al.*, 2020). This is likely due to the long-read sequencing's ability to "anchor" the repetitive sequences on the two sides of the repeats due to their length, thus avoiding collapsing those into shorter regions.

Telomer sequences were identified on both ends of 9 chromosomes, showing their completeness. Interestingly, in some cases we identified telomeric repeat sequences inside chromosomes as well (**Figure 3.**). The presence of these has been showed earlier by FISH method (Forsyth *et al.*, 2005), and it is similar to what we observed in the case of the brown hare as well (Michell *et al.*, 2024b).

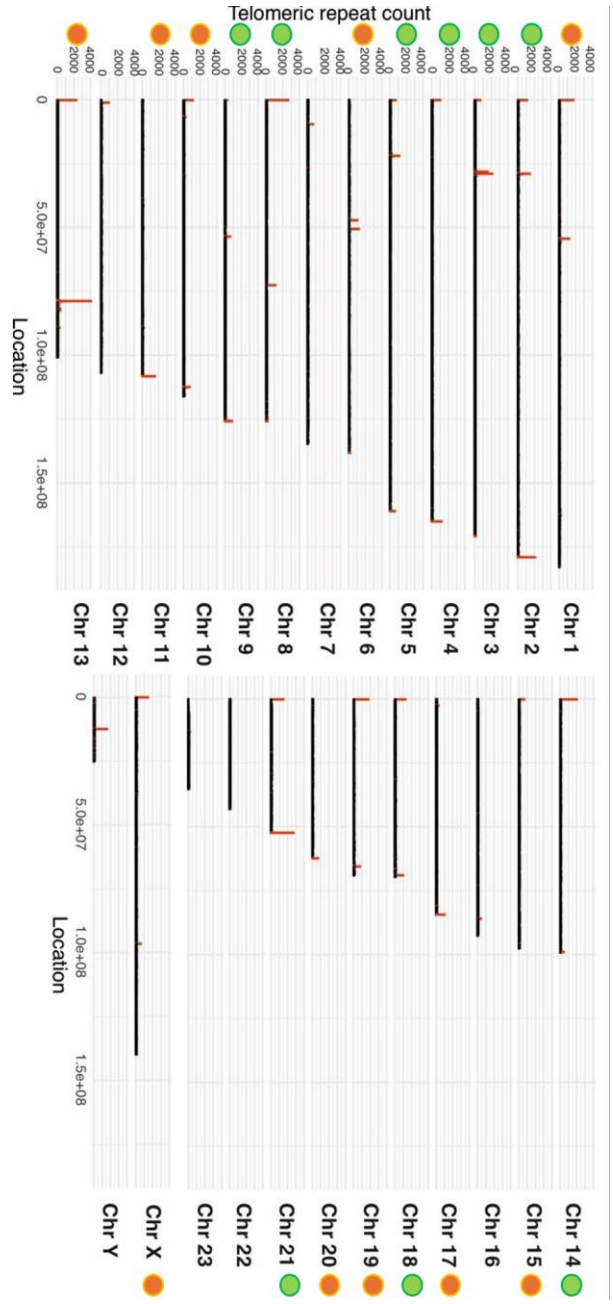

Figure 3 *– Telomer sequences along each chromosome: telomer sequences are marked by red peaks along the black lines symbolizing the chromosomes. The height of the peaks is relative to the amount of telomeric repeats detected at each location. Green dots mark chromosomes with telomeric repeats identified at both ends (chromosomes 2, 3, 4, 5, 8, 9, 14, 18 and 21), while orange dots mark chromosomes with telomeres identified only on one end (chromosomes 1, 6, 10, 11, 13, 15, 17, 19, 20 and X)*

Comparing the genome assemblies of the mountain hare and the brown hare, we found high synteny between the two high quality assemblies. Compared to this, our mountain hare assembly showed a lower degree of synteny with the mountain hare draft genome (**Figure 4.**), which is explained by the latter having been scaffolded based on the European rabbit's (*Oryctolagus cuniculus*) OryCun2 (NCBI GenBank accession: GCA_000003625.1) reference genome (Marques *et al.*, 2020). The nucleotide level sequence similarity was as expected, with the two mountain hare assemblies having a greater similarity to each other, than to the brown hare.

Overall, we were able to assign 99,75% of the full sequence length to either one of the 23 autosomes or the X or Y chromosomes. The BUSCO evaluation found 95,1% of the search orthologs of the mammalia_odb10 (9226 orthologs in total), and 93,2% of the glires_odb10 orthologs (13 798 total). The assembled mitochondrial DNA sequence was 17 482 bases, circular, and we were able to annotate all expected genes on it.

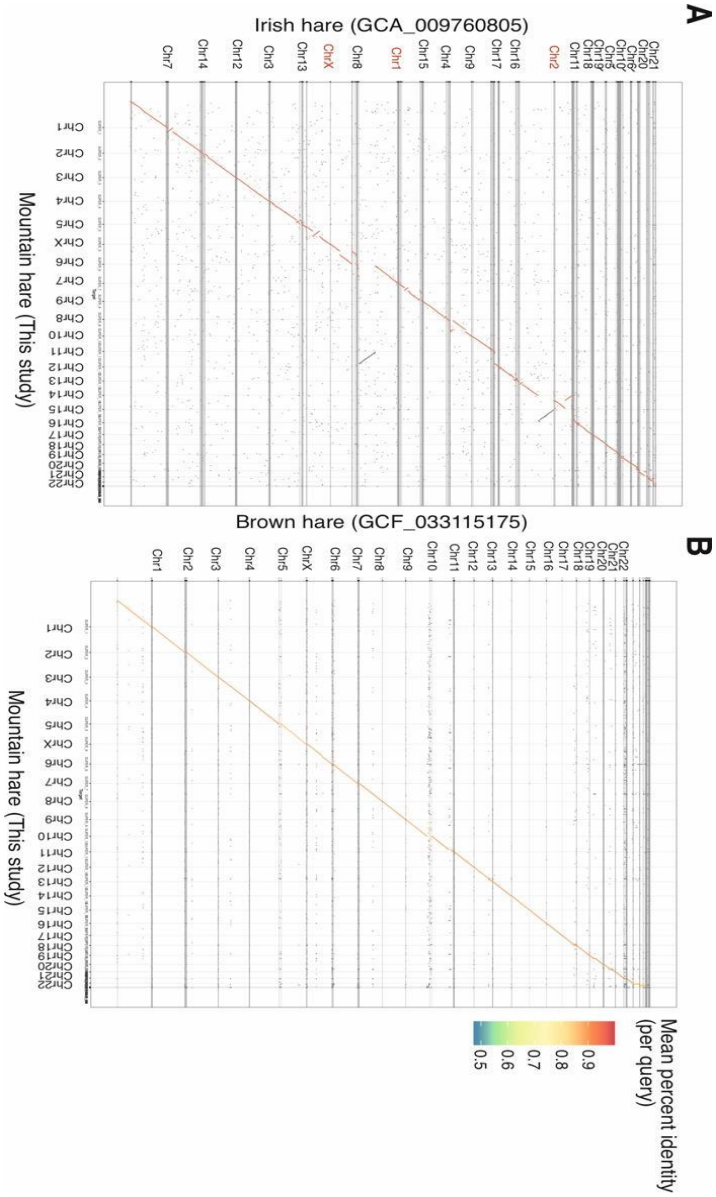

Figure 4- *Comparison of the newly assembled mountain hare (Lepus timidus timidus, "Mountain hare (this study)") with the Irish hare (Lepus timidus hibernicus) and B) brown hare (Lepus europaeus) genomes. The Irish hare genome was scaffolded based on the European rabbit, and the difference between the chromosome numbers and organization are to visible in the chromosomes marked in red (Chr2, Chr1, ChrX), while nucleotide level sequence similarity remains high (mostly over 90%), as expected in the case of subspecies. Contrary to this, when comparing to the brown hare genome, nucleotide sequence similarity is lower, while chromosome synteny is higher due to the de novo scaffolding in both cases. (Fekete et al., 2025a)*

3.3. Mitochondrial DNA recombination in mammalian cells

Mitochondrial DNA was isolated from heteroplasmic MGME1 knockout cell lines, which increases the chance of recombination and its detection. Despite the enrichment for mitochondrial DNA before sequencing, the resulting data still contained a high amount of nuclear sequences, while also showing distinct mitochondrial peaks in the diagrams of GC content and sequence read length distribution (**Figure 5.**).

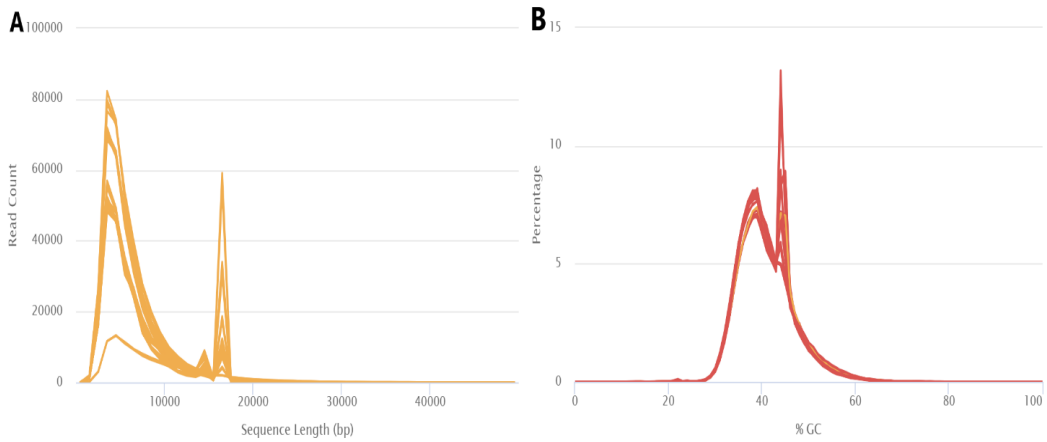

Figure 5 *- A) Read length and count of mitochondrial DNA sequencing. There is a large peak at around the length of the complete mitochondrial genome, between 15 – 20 kbp. B) GC% distribution of the reads. Beside the near normal distribution of nuclear genomic reads, we see a large peak at the values typical for mtDNA.*

As a first step, chimeric alignment between mitochondrial and nuclear sequences were removed, same as purely nuclear sequences. Purely mitochondrial sequences were re-aligned to haplotype specific references. Following this, we identified 24 distinctive SNP-s and a large indel in the BH10, and 89 distinctive SNPs in the HP202B cell line. Recombinant DNA molecules were identified at read level using these variants and scripts written by me for this goal (Zenodo: https://zenodo.org/records/10473729) With the applied method and settings, the rate of recombination reads was 12% (809 recombinant of 6730 pure mtDNA

read) in the BH10 and 0,8% (97 recombinant of 11 976 pure mtDNA read) in the HP202B cell lines. A visual example of recombinant and non-recombinant reads is shown in **Figure 6**.

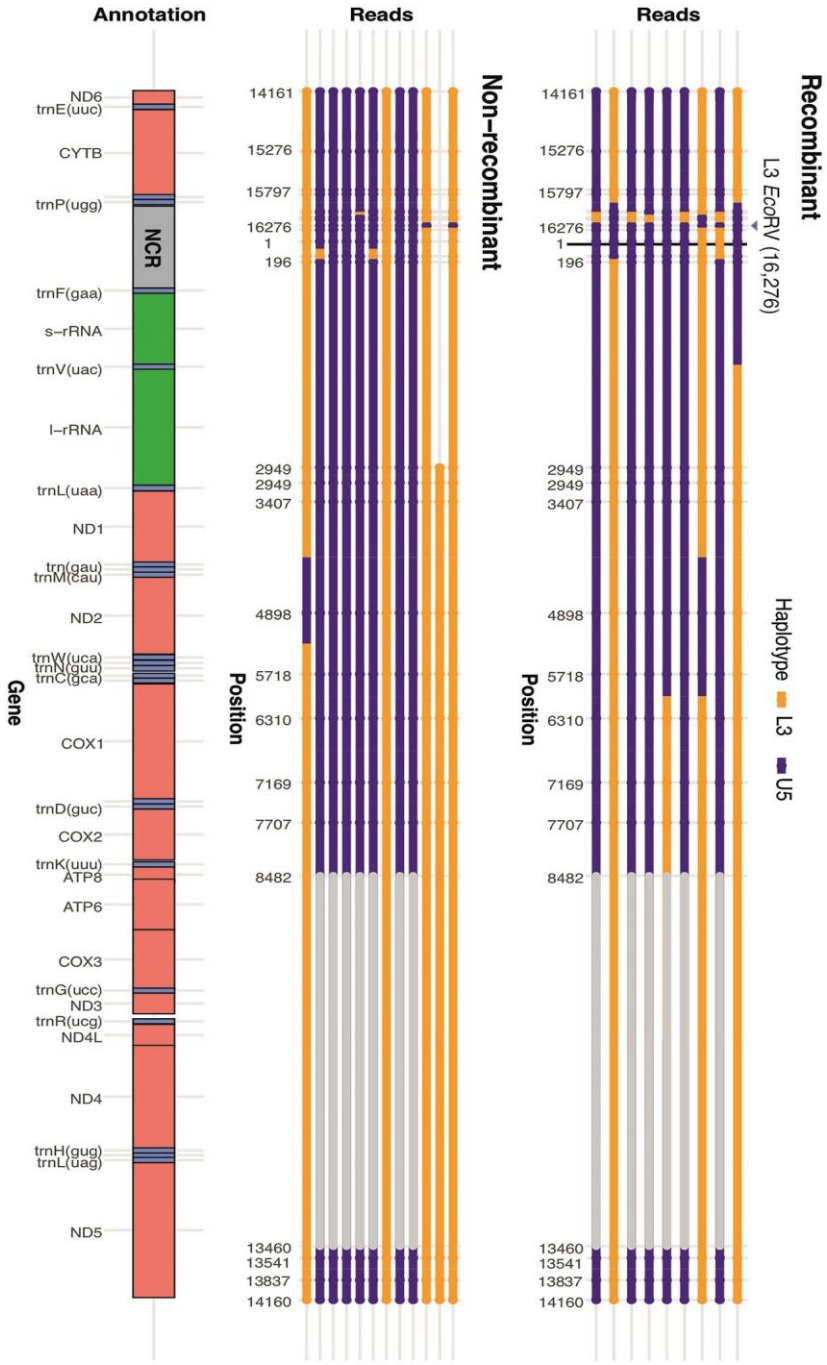

Figure 5 *– Sequences identified as recombinant and non-recombinant from the BH10 cell line (Fragkoulis et al., 2024). Differing colors in the sequences marked as non-recombinant signify segments that were identified as belonging to the opposing haplotype, yet not considered recombinant due to low confidence.*

4. Conclusions and suggestions

- Even though the free-living Hungarian rabbit population (Tiszaszentimre) we examined showed no signs of admixture with domestic rabbits, it is worth to consider further sampling across the country, as we know little of the state of rabbit populations in Hungary.
- The functions of intergenic and other non-coding regions of the genome are not well known, but the enrichment of certain transcription factor binding sites in putatively selected regions and some of the selected regions not overlapping with any known genes, supports the importance of these in domestication and selection.
- These experiments could be further expanded, and also, their accuracy could be enhanced by the use of a newly available, highly continuous and complete reference genome.
- The newly assembled, chromosome-level mountain hare (*Lepus timidus*) genome is currently listed as the reference genome of the species by the NCBI (GenBank: GCA_040893245.2).
- This *Lepus timidus* assembly, together with the *Lepus europaeus* assembly generated by our research group, is currently being utilized in further genomic research.
- Using a new bioinformatic method based on highly accurate long-read sequencing I could confirm the possibility of the recombination of mtDNA in mammalian cell lines. My method identifies recombinant molecules at read level.
- Due to the possibility of mtDNA recombination, the possibility of evolutionary significance of it cannot be ignored either. Especially in cases, where heteroplasmy and hybridization between populations is prevalent, and where anomalies can be possibly explained by it.

5. New scientific achievements

1. We have produced and made public a large amount (42) of high coverage (~35x) whole genome sequencing data of Hungarian rabbits, both of wild rabbit (*Oryctolagus cuniculus cuniculus*) populations and domestic rabbit breeds (*Oryctolagus cuniculus domesticus*). These populations did not have a comparable amount of genomic information available before.
2. Based on genomic data, we have shown that some European rabbit populations of Hungary (*Oryctolagus cuniculus cuniculus)* show signs of admixture with domestic rabbits (*Oryctolagus cuniculus domesticus)*, using the ADMIXTURE and Treemix softwares as well as D and f4 statistics.
3. I have identified 46 genomic regions that have been under selection during the domestication process of the rabbit. For this, I used three methods combined: Composite Likelihood Ratio, Fst and nucleotide diversity.
4. I have found enrichment of certain transcription factor binding site (TFBS) types in the above mentioned selected regions, that possibly have played a role in the domestication process of the European rabbit. Overall 7 types of TFBS have shown significant enrichment. Of these, 1170 individual sites were found in the selected regions, 93 of these in regions not overlapping with genes.
5. We have sequenced and assembled the genome of the mountain hare (Lepus timidus timidus) to chromosome level, up to current standards in methodology and end results. The resulting high quality genome is publicly available, and currently serves as a primary reference genome of the species in the NCBI assembly database.
6. I have developed a bioinformatics method which uses read level identification of recombinant sequences to support the long contested possibility of mitochondrial DNA recombination in mammalian cells.

6. Publications on the topic of the thesis

Fekete, Z., Német, Z., Ninausz, N., Fehér, P., Schiller, M., Alnajjar, M., Szenes, Á., Nagy, T., Stéger, V., Kontra, L., és Barta, E. (2025). Whole-Genome Sequencing-Based Population Genetic Analysis of Wild and Domestic Rabbit Breeds. *Animals*, *15*(6), 775. https://doi.org/10.3390/ani15060775, **IF: 2.7, SJR: 0.733, Q1**

Fekete, Z., Absolon, D. E., Michell, C., Wood, J. M. D., Goffart, S., és Pohjoismäki, J. L. O. (2025). Chromosome-level reference genome assembly for the mountain hare (Lepus timidus). *Peer Community Journal*, *5*, e14. https://doi.org/10.24072/pcjournal.514, **SJR: 0.797, Q1**

Fragkoulis, G., Hangas, A., **Fekete, Z.,** Michell, C., Moraes, C. T., Willcox, S., Griffith, J. D., Goffart, S., és Pohjoismäki, J. L. O. (2024). Linear DNA-driven recombination in mammalian mitochondria. *Nucleic Acids Research*, *52*(6), 3088–3105. https://doi.org/10.1093/nar/gkae040, **IF: 13.1, SJR: 7.776, Q1**

Michell, C., Collins, J., Laine, P. K., **Fekete, Z.,** Tapanainen, R., Wood, J. M. D., Goffart, S., és Pohjoismäki, J. L. O. (2024). High quality genome assembly of the brown hare (Lepus europaeus) with chromosome-level scaffolding. *Peer Community Journal*, *4*, e26. https://doi.org/10.24072/pcjournal.393, **SJR: 0.797, Q1**

Tapanainen, R., Aasumets, K., **Fekete, Z.,** Goffart, S., Dufour, E., és L. O. Pohjoismäki, J. (2024). Species-specific variation in mitochondrial genome tandem repeat polymorphisms in hares (Lepus spp., Lagomorpha, Leporidae) provides insight into their evolution. *Gene*, *926*, 148644. https://doi.org/10.1016/j.gene.2024.148644, **IF: 2.4, SJR 0.682, Q2**

7. References

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Auwera, G. A. V. der, & O'Connor, B. D. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. O'Reilly Media, Incorporated.

Bohling, J. (2020). Evaluating the effect of reference genome divergence on the analysis of empirical RADseq datasets. *Ecology and Evolution*, *10*(14), 7585–7601. https://doi.org/10.1002/ece3.6483

Brown, M. R., Manuel Gonzalez De La Rosa, P., & Blaxter, M. (2025). tidk: A toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics*, *41*(2), btaf049. https://doi.org/10.1093/bioinformatics/btaf049

Carneiro, M., Rubin, C.-J., Di Palma, F., Albert, F. W., Alföldi, J., Martinez Barrio, A., Pielberg, G., Rafati, N., Sayyab, S., Turner-Maier, J., Younis, S., Afonso, S., Aken, B., Alves, J. M., Barrell, D., Bolet, G., Boucher, S., Burbano, H. A., Campos, R., … Andersson, L. (2014). Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science (New York, N.Y.)*, *345*(6200), 1074–1079. https://doi.org/10.1126/science.1253714

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, *18*(2), 170–175. https://doi.org/10.1038/s41592-020-01056-5

Czipa, E., Schiller, M., Nagy, T., Kontra, L., Steiner, L., Koller, J., Pálné-Szén, O., & Barta, E. (2020). ChIPSummitDB: A ChIP-seq-based database of human transcription factor binding sites and the topological arrangements of the proteins bound to them. *Database: The Journal of Biological Databases and Curation*, *2020*, baz141. https://doi.org/10.1093/database/baz141

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics (Oxford, England)*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

DeGiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I., & Nielsen, R. (2016). SweepFinder2: Increased sensitivity, robustness and flexibility. *Bioinformatics (Oxford, England)*, *32*(12), 1895–1897. https://doi.org/10.1093/bioinformatics/btw051

Fekete, Z., Absolon, D. E., Michell, C., Wood, J. M. D., Goffart, S., & Pohjoismäki, J. L. O. (2025). Chromosome-level reference genome assembly for the mountain hare (Lepus timidus). *Peer Community Journal*, *5*, e14. https://doi.org/10.24072/pcjournal.514

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, *117*(17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

Formenti, G., Abueg, L., Brajuka, A., Brajuka, N., Gallardo-Alba, C., Giani, A., Fedrigo, O., & Jarvis, E. D. (2022). Gfastats: Conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics*, *38*(17), 4214–4216. https://doi.org/10.1093/bioinformatics/btac460

Forsyth, N. R., Elder, F. F. B., Shay, J. W., & Wright, W. E. (2005). Lagomorphs (rabbits, pikas and hares) do not use telomere-directed replicative aging in vitro. *Mechanisms of Ageing and Development*, *126*(6–7), 685–691. https://doi.org/10.1016/j.mad.2005.01.003

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hinrichs, A. S. (2006). The UCSC Genome Browser Database: Update 2006. *Nucleic Acids Research*, *34*(90001), D590–D598. https://doi.org/10.1093/nar/gkj144

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., Davidson, C., Dodiya, K., El Houdaigui, B., Fatima, R., … Flicek, P. (2021). Ensembl 2021. *Nucleic Acids Research*, *49*(D1), D884–D891. https://doi.org/10.1093/nar/gkaa942

Kraytsberg, Y., Schwartz, M., Brown, T. A., Ebralidse, K., Kunz, W. S., Clayton, D. A., Vissing, J., & Khrapko, K. (2004). Recombination of Human Mitochondrial DNA. *Science*, *304*(5673), 981–981. https://doi.org/10.1126/science.1096342

Ladoukakis, E. D., & Zouros, E. (2001). Direct Evidence for Homologous Recombination in Mussel (Mytilus galloprovincialis) Mitochondrial DNA. *Molecular Biology and Evolution*, *18*(7), 1168–1175. https://doi.org/10.1093/oxfordjournals.molbev.a003904

Lee, T.-H., Guo, H., Wang, X., Kim, C., & Paterson, A. H. (2014). SNPhylo: A pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics*, *15*, 162. https://doi.org/10.1186/1471-2164-15-162

Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: Assessing Genomic Data Quality and Beyond. *Current Protocols*, *1*(12), e323. https://doi.org/10.1002/cpz1.323

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 Biology and Evolution*, *12*(1), 3656–3662. https://doi.org/10.1093/gbe/evz273

Meng, G., Li, Y., Yang, C., & Liu, S. (2019). MitoZ: A toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Research*, *47*(11), e63–e63. https://doi.org/10.1093/nar/gkz173

Michell, C., Collins, J., Laine, P. K., Fekete, Z., Tapanainen, R., Wood, J. M. D., Goffart, S., & Pohjoismäki, J. L. O. (2024). High quality genome assembly of the brown hare (Lepus europaeus) with chromosome-level scaffolding. *Peer Community Journal*, *4*, e26. https://doi.org/10.24072/pcjournal.393

Nevado, B., Ramos-Onsins, S. E., & Perez-Enciso, M. (2014). Resequencing studies of nonmodel organisms using closely related reference genomes: Optimal experimental designs and bioinformatics approaches for population genomics. *Molecular Ecology*, *23*(7), 1764–1779. https://doi.org/10.1111/mec.12693

Peeva, V., Blei, D., Trombly, G., Corsi, S., Szukszto, M. J., Rebelo-Guiomar, P., Gammage, P. A., Kudin, A. P., Becker, C., Altmüller, J., Minczuk, M., Zsurka, G., & Kunz, W. S. (2018). Linear mitochondrial DNA is rapidly degraded by components of the replication machinery. *Nature Communications*, *9*(1), 1727. https://doi.org/10.1038/s41467-018-04131-w

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*(11), e1002967. https://doi.org/10.1371/journal.pgen.1002967

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Lee, C., Ko, B. J., Kim, J., Bista, I., Smith, M., Haase, B., … Jarvis, E. D. (2020). *Towards complete and error-free genome assemblies of all vertebrate species*. https://doi.org/10.1101/2020.05.22.110833

Sheng, P., Hu, Y., Sun, Z., Yang, L., Hu, S., Fuller, B. T., & Shang, X. (2020). Early commensal interaction between humans and hares in Neolithic northern China. *Antiquity*, *94*(375), 622–636. https://doi.org/10.15184/aqy.2020.36

Smit, A., Hubley, R., & Green, P. (2015). *RepeatMasker Open-4.0* [Computer software]. repeatmasker.org

Somerville, A. D., Sugiyama, N., Manzanilla, L. R., & Schoeninger, M. J. (2017). Leporid management and specialized food production at Teotihuacan: Stable isotope data from cottontail and jackrabbit bone collagen. *Archaeological and Anthropological Sciences*, *9*(1), 83–97. https://doi.org/10.1007/s12520-016-0420-2

Tapanainen, R., Aasumets, K., Fekete, Z., Goffart, S., Dufour, E., & L. O. Pohjoismäki, J. (2024). Species-specific variation in mitochondrial genome tandem repeat polymorphisms in hares (Lepus spp., Lagomorpha, Leporidae) provides insight into their evolution. *Gene*, *926*, 148644. https://doi.org/10.1016/j.gene.2024.148644

Van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, *30*(2), E386–E394. https://doi.org/10.1002/humu.20921

Vinogradov, A. E. (1998). Genome size and GC-percent in vertebrates as determined by flow cytometry: The triangular relationship. *Cytometry*, *31*(2), 100–109. https://doi.org/10.1002/(SICI)1097-0320(19980201)31:2<100::AID-CYTO5>3.0.CO;2-Q

Zhou, C., McCarthy, S. A., & Durbin, R. (2023). YaHS: Yet another Hi-C scaffolding tool. *Bioinformatics*, *39*(1), btac808. https://doi.org/10.1093/bioinformatics/btac808

Zsurka, G., Hampel, K. G., Kudina, T., Kornblum, C., Kraytsberg, Y., Elger, C. E., Khrapko, K., & Kunz, W. S. (2007). Inheritance of Mitochondrial DNA Recombinants in Double-Heteroplasmic Families: Potential Implications for Phylogenetic Analysis. *The American Journal of Human Genetics*, *80*(2), 298–305. https://doi.org/10.1086/511282

Zsurka, G., Kraytsberg, Y., Kudina, T., Kornblum, C., Elger, C. E., Khrapko, K., & Kunz, W. S. (2005). Recombination of mitochondrial DNA in skeletal muscle of individuals with multiple mitochondrial DNA heteroplasmy. *Nature Genetics*, *37*(8), 873–877. https://doi.org/10.1038/ng1606