

DOKTORI (Ph.D.) ÉRTEKEZÉS TÉZISEI

FEKETE ZSÓFIA

GÖDÖLLŐ

2025



Magyar Agrár- és Élettudományi Egyetem

GENOMIKAI VIZSGÁLATOK BIOINFORMATIKAI
MÓDSZEREKKEL LEPORIDAE FAJOKBAN ÉS
MITOKONDRIÁLIS SZEKVENCIÁKON

Fekete Zsófia

2025.

Gödöllő

A doktori iskola

Megnevezése: **Állatbiotechnológiai és Állattudományi Doktori Iskola**

Tudományága: Állatbiotechnológia

Vezetője: **Prof. Dr. Mézes Miklós**

Az MTA rendes tagja, DSc, Egyetemi tanár; Tanszékvezető

Magyar Agrár- és Élettudományi Egyetem

Élettani és Takarmányozástani Intézet, Takarmánybiztonsági Tanszék

Témavezetők: **Dr. Kontra Levente**

PhD, Központi egység vezető

HUN-REN; Kísérleti Orvostudományi Kutatóintézet

Bioinformatikai Központi Egység

Dr. Hoffmann Orsolya Ivett

PhD, Tudományos főmunkatárs, csoportvezető

Magyar Agrár- és Élettudományi Egyetem

Genetika és Biotechnológia Intézet, Állatbiotechnológia Tanszék

.....
Prof. Dr. Mézes Miklós

Iskolavezető

.....
Dr. Kontra Levente

Témavezető

.....
Dr. Hoffmann Orsolya Ivett

Témavezető

Tartalomjegyzék

1. A munka előzményei és célkitűzéseim.....	2
1.1. Célkitűzések	3
1.1.1. Populációgenomikai vizsgálatok üregi nyúlban.....	3
1.1.2. Havasi nyúl (<i>Lepus timidus</i>) referenciagenom összeszerelése.....	3
1.1.3. Lineáris DNS-vezérelt rekombináció igazolása mitokondriális szekvenciákon bioinformatikai módszerrel.....	4
2. Anyag és módszer.....	4
2.1. Teljes genom szekvenálás alapú populációgenomikai vizsgálatok üregi- és házinyúlban	4
2.2. <i>Lepus timidus</i> referenciagenom összeszerelése.....	5
2.3. Mitokondriális DNS rekombinációjának igazolása emlős sejtekben	6
3. Eredmények és diszkusszió.....	7
3.1. Hibridizáció és szelekció az üregi- és házinyúl között	7
3.2. Havasi nyúl (<i>Lepus timidus</i>) referenciagenom.....	11
3.3. Mitokondriális DNS rekombinációja emlősökben	15
4. Következtetések és javaslatok.....	18
5. Új tudományos eredmények.....	19
6. Publikációk az értekezés témakörében.....	20
1. Az értekezés témakörében megjelent tudományos cikkek.....	20
2. Konferenciaközlemények az értekezés témakörében... Error! Bookmark not defined.	
7. Hivatkozások.....	22

1.A munka előzményei és célkitűzéseim

A *Leporidae* fajok a világ minden táján elterjedtek, változatos élőhelyeken és szerepekben találkozhatunk velük. Az üregi nyúl (*Oryctolagus cuniculus*) az egyetlen ma élő *Oryctolagus* faj, amelynek azonban létezik vadon élő és házasított változata is. Ezzel szemben a *Lepus* nemzetség számos ma élő fajának képviselői világszerte megtalálhatóak. A nemzetség széles körű elterjedésének, és bizonyos esetekben emberrel való asszociációjuknak (Sheng *et al.*, 2020; Somerville *et al.*, 2017) ellenére azonban nem kerültek házasításra.

A populációgenomikai vizsgálatok, ezen belül is a szelekció vizsgálata állattenyésztési, evolúciós és ökológiai szempontból is jelentős. A modern szekvenálási technológiák felhasználásával nagy pontossággal és felbontással vizsgálható a szelekció, hibridizáció és introgresszió. A mesterséges és a természetes szelekció és hibridizáció genomi lenyomatai egymáshoz hasonlóak, így hasonló módszerekkel vizsgálhatóak. A vizsgálatokban jelentős szerepe van a referenciagenomoknak, és ezek minőségének. Az alkalmazott referencia befolyásolja az illesztést, és ennek következtében a hívott változatokat, és a további vizsgálatok eredményét (Bohling, 2020; Nevado *et al.*, 2014). Az úgynevezett draft genomok leváltása nagy folytonosságú, kromoszóma szintű összeszerelésekkel fontos az ilyen hatások minimalizálásának érdekében.

A mitokondriális markerek elterjedten alkalmazottak filogenetikai és populációgenomikai vizsgálatokban, ugyanakkor hibridizálódó populációk között gyakori az mtDNS introgressziója. Az introgresszió potenciálisan szelekciós előnyt jelenthet a fogadó fajnak, emellett azonban zavaró tényezőként jelenik meg evolúciós vizsgálatokban. Számos fajban ismert a mitokondriális DNS rekombinációjának lehetősége, egyes esetben evolúciós jelentősége is (Ladoukakis és Zouros, 2001). Bár a mitokondriális DNS emlősökben maternális öröklődésű, és emiatt a rekombináció, ha végbe is megy, a legtöbb esetben nem okoz változást, a heteroplazmia előfordulása ismert. Aktívan hibridizálódó fajok

között – például *Lepus timidus* és *Lepus europaeus* (Tapanainen *et al.*, 2024) esetében – nem kizárható az esetleges, evolúciós szempontból jelentős rekombináció sem.

1.1. Célkitűzések

Munkám céljával tűztem ki szelekciós és rekombinációs folyamatok vizsgálatát modern szekvenálási technológiák megfelelő kombinációjával emlős fajokban, illetve a jövőben tervezett vizsgálatok háttérének megteremtését.

1.1.1. Populációgenomikai vizsgálatok üregi nyúlban

Az üregi nyúlról rendelkezésre álló nagy mennyiségű információ ellenére keveset tudunk a közép-európai, így a magyarországi vadon élő üregi nyúl populációkról, azok eredetéről, illetve pillanatnyi állapotáról. Célul tűztem ki a vadon élő üregi nyulak helyzetének vizsgálatát Magyarországon, beleértve az esetleges kereszteződést házi nyulakkal és a beltenyésztettség mértékét.

Célom ezen felül a háziasítás során történt szelekció vizsgálata. Dolgozatomban megvizsgáltam a szelektált régiók és változatok jelenlétét és ezek tulajdonságait, ideértve mennyiségüket, elhelyezkedésüket a genomon, és feltételezhető szerepüket.

A vizsgálatok elvégzéséhez publikus, és saját teljes genom szekvenálási adatokat is felhasználtam. Az általunk szekvenált magyar üregi nyúl és házinyúl minták mellett az NCBI SRA adatbázisából gyűjtött ibériai és franciaországi, publikusan elérhető házi- és vadon élő üregi nyulakból származó minták integrációjával szerettünk volna teljesebb képet elérni.

*1.1.2. Havasi nyúl (*Lepus timidus*) referenciagenom összeszerelése*

Célul tűztem ki egy, eddig referenciagenommal nem rendelkező *Leporidae* faj (havasi nyúl, *Lepus timidus*) genomjának modern sztenderdeknek megfelelő, referencia minőségű összeszerelését a pillanatnyilag elfogadott

legjobb gyakorlatot követve, és az összeszerelt genom összevetését egy hibridizálódó közeli fajnak, a mezei nyúlnek (*Lepus europaeus*) genomjával.

1.1.3. Lineáris DNS-vezérelt rekombináció igazolása mitokondriális szekvenciákon bioinformatikai módszerrel

Célom volt bioinformatikai módszer fejlesztése rekombináció detektálására hosszú leolvasásos (PacBio HiFi, Wenger *et al.*, 2019) szekvenálásból, mitokondriális szekvenciákon.

A rekombináció detektálására létező számos módszer általában a meiotikus rekombináció azonosítására törekszik, illetve olyan prokarióta- és virális genomokat vizsgál statisztikai alapon, amelyeknél a rekombináció léte széles körben ismert és elfogadott. Annak ellenére, hogy egyes heteroplazmikus esetekben korábban már detektáltak rekombináns emlős mtDNS-t (Kraytsberg *et al.*, 2004; Zsurka *et al.*, 2005, 2007).

Az általam fejlesztett módszer célja a rekombináció előfordulásának bizonyítása emlős sejtekben, mitokondriális szekvenciákon, és ez által laboratóriumi eredmények bioinformatikai hitelesítése.

2. Anyag és módszer

2.1. Teljes genom szekvenálás alapú populációgenomikai vizsgálatok üregi- és házinyúlban

100 üregi- és házinyúl teljes genom szekvenálást használtunk fel, amelyben az NCBI SRA-ról gyűjtött publikus szekvenálási adatok és általunk izolált és szekvenált minták is szerepeltek. Az üregi nyúl minták között magyar – saját minták –, franciaországi és ibériai-félszigeti – publikus minták – szerepeltek, emellett változatos házinyúl fajtákból származó mintákat használtunk fel. A munkámhoz izolált minták nagy-lefedettséggel kerültek teljes genom szekvenálással. A publikus minták lefedettsége alacsonyabb volt, de a GATK ToolKit (Auwera és O'Connor, 2020) ajánlott jó gyakorlatát követve olyan, jó

minőségű variáció szettet kaptunk, amelyben a minőségi szűrések után a vizsgálatokhoz használt pontok kivétel nélkül minden mintában genotipizáltak voltak.

A populációstruktúra vizsgálatát az SNP-k felhasználásával az SNPhylo (Lee *et al.*, 2014) pipeline -filogenetikai fa készítés – és a plink1.9 (Chang *et al.*, 2015) – főkomponens analízis – program segítségével végeztem. Megvizsgáltam az üregi- és házinyulak közötti hibridizáció lehetőségét ADMIXTURE (Alexander *et al.*, 2009) és Treemix (Pickrell és Pritchard, 2012) programokkal, további statisztikákkal, mint a D és f₄ alátámasztva. Az egyes, egyedileg szekvenált genomokon található homozigóta szakaszok hossza alapján vizsgáltam a beltenyészettségi mértékét is, a plink 1.9 használatával.

A domesztikáció hatását, és a házinyulakban szelektált genomi szakaszokat a potenciálisan hibrid mintákat kizárva vizsgáltam. A szelekció vizsgálatát három módszer kombinációjával végeztem, a nukleotid diverzitás (π), a fixációs index (F_{st}) és a CLR (composite likelihood ratio) felhasználásával. Előbbi kettőhöz a vcftools (Danecek *et al.*, 2011), utóbbihoz a SweepFinder2 (DeGiorgio *et al.*, 2016) szoftvert használtam. Az eredmények kiértékelésekor az adódott extrém értékeket (99. percentilis) vettem figyelembe. A három módszerrel potenciálisan szelektáltak adódott régiókat összevettem egymással, majd az átfedő régiókat az azokban található gének és transzkripciós faktor kötőhelyek (TFBS) szempontjából vizsgáltam. A gének azonosítását az Ensembl (release 104) (Howe *et al.*, 2021) annotációja alapján végeztem, TFBS-ek meghatározását pedig LiftOverrel (Hinrichs, 2006), a ChipSummitDB (Czipo *et al.*, 2020) adatbázis alapján.

2.2. *Lepus timidus* referenciagenom összeszerelése

A genom-összeszerelés alapjául egy hím *Lepus timidus* egyed szolgált, amely Finnországban a vadászati szezon alatt került kilövésre. A fajazonosság és az állat nem-hibrid mivolta morfológiai jegyek és genetikai markerek alapján is

megerősítésre került. Az állatból létrehozott fibroblaszt sejtvonalból került izolálásra a DNS a PacBio HiFi és az Illumina Hi-C szekvenáláshoz is. A mitokondriális DNS izolálása ugyanebből az egyedből történt, de a genomi DNS-től eltérően fűlszövetből. A mtDNS PCR amplifikációt követően Illumina módszerrel volt szekvenálva.

A genom összeszerelésekor a modern sztenderdeket (Rhie *et al.*, 2020) követtem. A minőségi kontrollt követően Hifiasm (Cheng *et al.*, 2021) programmal szereltem össze a contigokat, majd a duplikációk eltávolítását követte a Hi-C scaffolding YaHS (Zhou *et al.*, 2023) szoftverrel, és a kézi kuráció. A repetitív szakaszok annotációját és maszkolását a RepeatModeler (Flynn *et al.*, 2020) és a RepeatMasker (Smit *et al.*, 2015) programokkal, a telomer szekvenciák azonosítását a tidk-val (Brown *et al.*, 2025) végeztem. A mitokondriális genomot a MitoZ (Meng *et al.*, 2019) pipeline-nal szereltem össze. A minőségi ellenőrzésekhez, illetve a teljesség becsléséhez a BUSCO (Manni *et al.*, 2021) – géntartalom, várt ortológok jelenléte és teljessége –, a gfastats (Formenti *et al.*, 2022) és QCAST (Gurevich *et al.*, 2013) – szekvencia tulajdonságok és statisztikák – programokat használtam.

2.3. Mitokondriális DNS rekombinációjának igazolása emlős sejtekben

A mitokondriális DNS rekombinációjának vizsgálatához két heteroplazmikus sejtvonalat használtunk, jól elkülöníthető haplotípusokkal. A haplocsoportok meghatározása a PhyloTree Build alapú webes elemző (Van Oven és Kayser, 2009) eszközzel történt. A felhasznált sejtvonalak egy eger (HP202B), és egy humán (BH10) heteroplazmikus vonal voltak. A rekombináció vizsgálatához MGME1 knockout vonalak kerültek felhasználásra, a lineáris mtDNS lebontásának (Peeva *et al.*, 2018) elnyomása, és az egyensúlynak a rekombináció irányába való eltolása érdekében. A mitokondriális DNS-t az izolálást követően PacBio HiFi technológiával szekvenáltuk.

A PacBio HiFi amplifikáció nélküli szekvenálási módszer, egy molekulát egy leolvasás reprezentál. A technológia nagy pontossága és az ismert haplotípusok segítségével a leolvasásokon következetesen megtalálható SNP-k és deléción alapján lehetséges volt azok haplotípusának azonosítása.

Tisztán mitokondriális leolvasásokat a leolvasások hosszúságát kihasználva választottam ki. A teljes egészükben mitokondriális szekvenciára illeszkedő, hosszukban a mitokondriális genommal közelítőleg megegyező leolvasások megbízhatóan valóban mitokondriális szekvenciák.

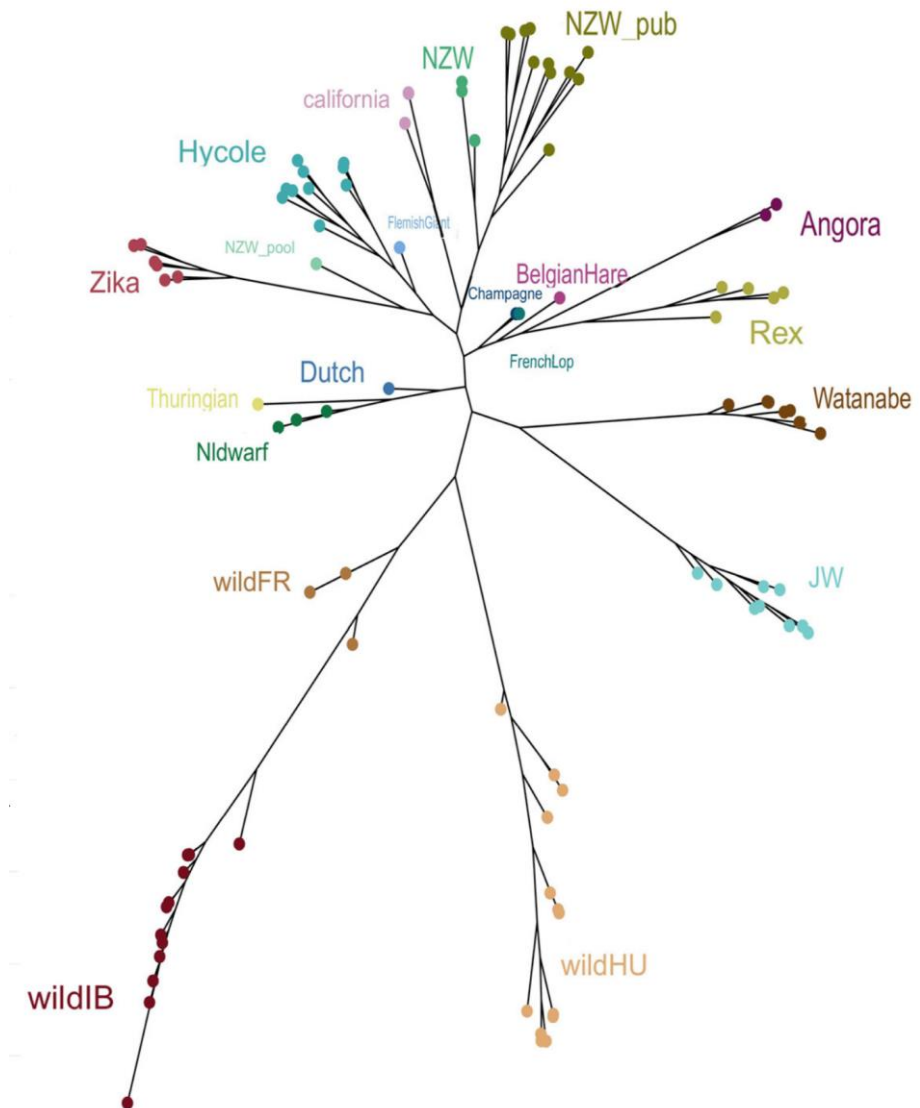
A haplotípusok azonosításához a leolvasásokat diploid genomról származóként kezelve heterozigótaként hívott variációkat használtam fel. A szekvenciákat az egyik jelenlévő haplotípusra specifikus referenciához illesztettem, a variációk hívását pedig a GATK ToolKit (Auwera és O'Connor, 2020) VariantCallerével és pbsv-vel (<https://github.com/PacificBiosciences/pbsv>) végeztem.

Az egyes leolvasásokon jelenlévő két haplotípushoz tartozó variációkból (A és B) pszeudoszekvenciát generáltam, kiértékeléséhez pedig egy pontozási rendszert állítottam fel. Valamennyi lépést párhuzamosan végeztem a két sejtvonagra, vizuális ellenőrzésekkel együtt. A célra írt scriptek elérhetők a <https://zenodo.org/records/10473729> címen.

3. Eredmények és diszkusszió

3.1. Hibridizáció és szelekció az üregi- és házinyúl között

A magyarországi üregi nyulak genetikailag elkülöníthetőnek adódtak az ibériai-félszigeti és franciaországi populációktól. A filogenetikai fán való elhelyezkedésük ezutóbbiakkal szomszédos, a házinyulaktól pedig távolabbi (**1. ábra**). Ez arra utal, hogy a genetikai különbségek természetes módon, a távolság és a környezet hatására alakultak ki, és a magyarországi üregi nyulak nem házinyulak elvadult leszármazottai.



1. ábra - Filogenetikai fa maximum-likelihood módszerrel. A házinyúl fajták jobbára a fajta célja szerint csoportosulnak, míg a vad üregi nyulak földrajzi eredet szerinti izolációt mutatnak. A két japán fajta itt is elkülönül a többi házinyúl csoporttól. Rövidítések: NZW – Új-Zélandi fehér; Nldwarf – Holland törpe, JW – Japán fehér, wildFR – franciaországi üregi nyúl; wildHU - magyarországi üregi nyúl; wildIB – ibériai-félszigeti üregi nyúl. (Fekete et al., 2025b)

A házi- és üregi nyúl hibridizációja több magyar üregi nyúl mintában is megfigyelhető volt, ezek mindegyike egyazon szubpopulációhoz tartozik. A keresztezés konzisztensnek adódott a különböző tesztek során, és a hibrid állatok mindegyike az állatkerti populációba tartozott (**2. ábra**). Ugyanakkor nem áll rendelkezésünkre információ arról, hogy honnan, hogyan és mikor történhetett pontosan a kereszteződés. A magyarországi üregi nyulak populációjáról összességében kevés információ áll rendelkezésre: az üregi nyulat célzó kutatások fókuszja őshazájukon, az Ibériai-félszigeten, vagy olyan helyeken van, ahol az invazív fajként van jelen, és kártevőként tartják számon.

A legtöbb, nyúl populációgenetikai és szelekciós kutatás a fajták közötti szelekcióval foglalkozik. A domesztikáció eredete sokgénés (Carneiro *et al.*, 2014), allélfrekvencia eltolódásokon alapul, nem pedig új mutációk megjelenésén és elterjedésén. Az üregi- és házinyúl közötti differenciációt és szelekciót három módszerrel megvizsgálva a szelektáltnak adódott szakaszokkal (összesen 46 szakasz) átfedő gének között egyedfejlődéssel, növekedéssel, jelátvitellel – viselkedéssel – és immunválasszal kapcsolatosak is adódtak. Ezek olyan tulajdonságokat befolyásolnak, amelyek változása általánosan jellemző a domesztikáció során.

A kiválasztott régiókban nagy mennyiségű erősen differenciált ($F_{st} > 0.75$, 983 variáció), de kevés teljesen fixált változat adódott. Emellett adódtak olyan szakaszok is (14), amelyeken nincs gén annotálva. Megvizsgáltuk a transzkripció faktor kötőhelyek (TFBS) jelenlétét is, és egyes típusok dúsulását tapasztaltuk a szelektált szakaszokban (Error! Reference source not found.), a szabályozó régiók fontosságát mutatva.

1. Táblázat - Szelektált régiókban azonosított, szignifikánsan dúsultnak adódó transzkripciós faktor kötőhelyek. Az első oszlop a genomon azonosított teljes mennyiséget, míg a harmadik az ebből szelektált régiókban található mutatót mutatja.

Összes	Név	Szelektált régióban	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

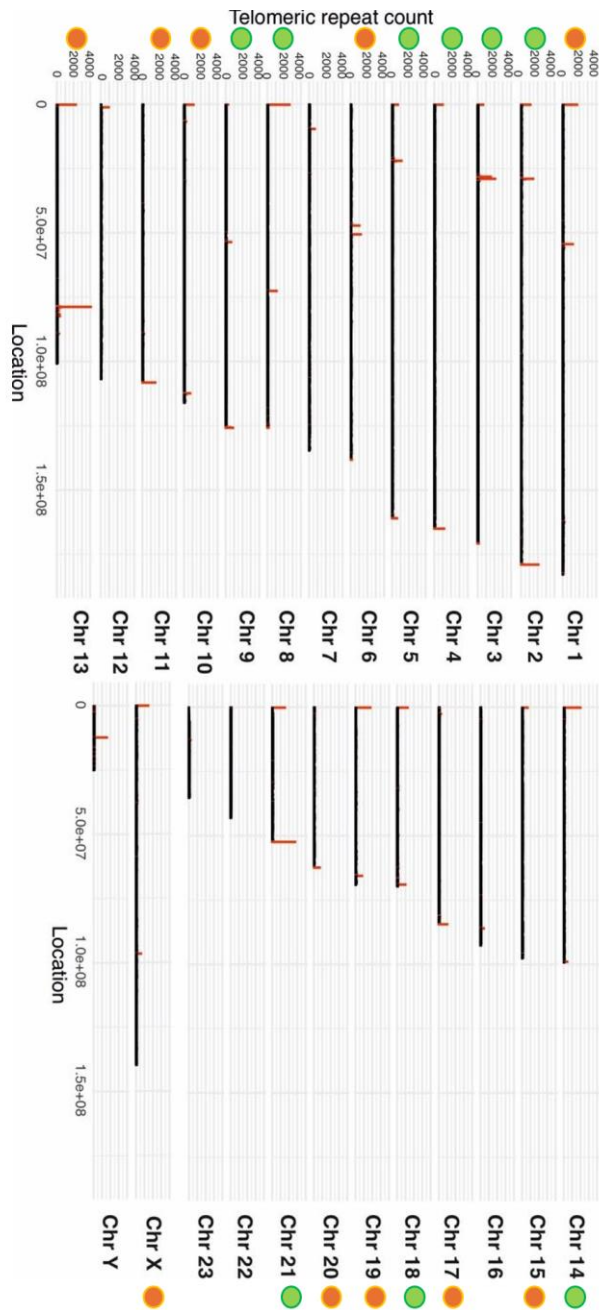
3.2. Havasi nyúl (*Lepus timidus*) referenciagenom

Nagy folytonosságú, kromoszómaszinten összeszerelt genomot hoztunk létre a havasi nyúlhoz (*Lepus timidus*), a mai sztenderdeknek megfelelően. A 2,7 Gbp hosszúságú genom körülbelül 500 Mbp-vel rövidebb az első, flowcitometrián alapuló becslésnél (Vinogradov, 1998), viszont nagyon közel áll viszont a *L. timidus* draft genomhoz (Marques *et al.*, 2020). A genomban nagy mennyiségű a repetitív szekvencia (42,35%) (Fekete *et al.*, 2025a), hasonlóan a mezei nyúléhoz (46%) (Michell *et al.*, 2024). Ez jelentősen többnek adódott a korábban becslésnél (23%) (Marques *et al.*, 2020). Ez valószínűleg annak köszönhető, hogy a hosszú leolvasásos szekvenálások jobban képesek megtartani a teljes repetitív szekvenciát azzal, hogy átérjük annak teljes hosszát, és „horgonyozzák” a két végén.

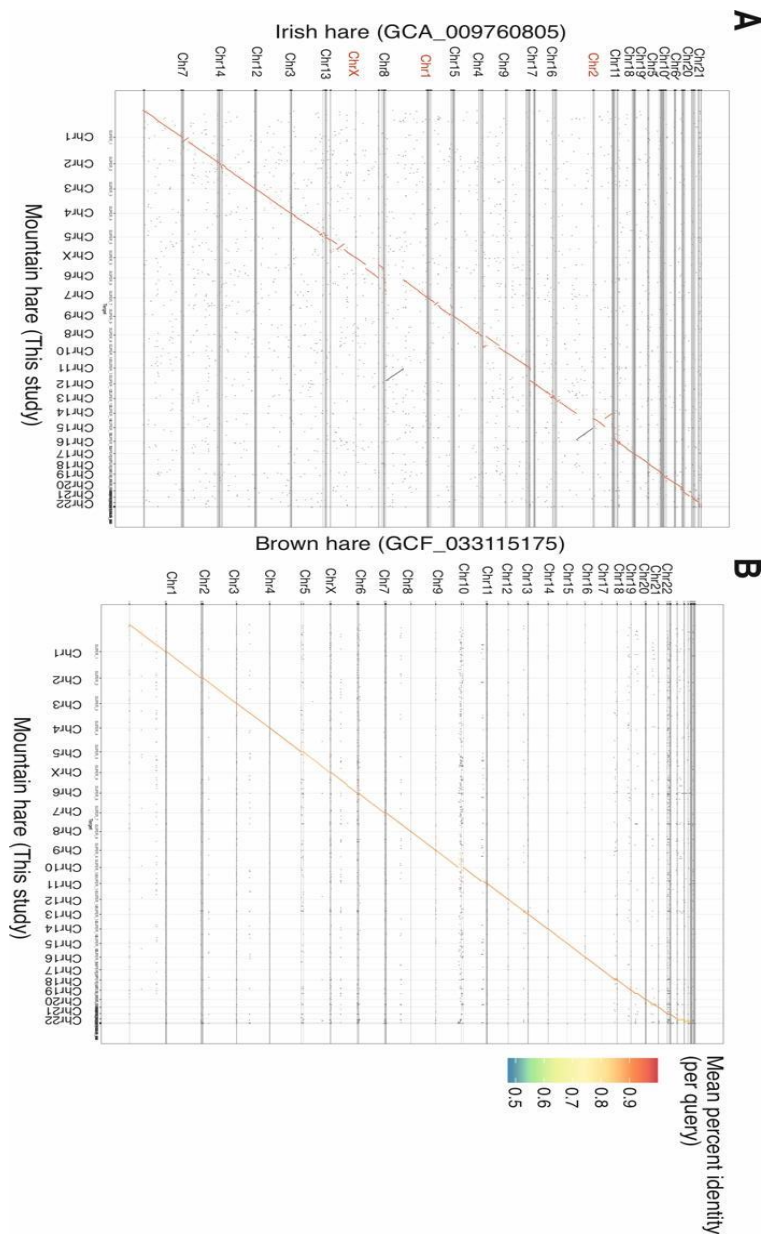
9 kromoszómának sikerült mindkét végén telomer szekvenciát azonosítani, azok teljességét mutatva. Érdekes módon egyes kromoszómákon belül is találhatóak telomer szakaszok (3. *ábra*). Ezek korábbi, FISH alapú megfigyelésekkel (Forsyth *et al.*, 2005) egybevágóak, és hasonlóak a mezei nyúlban (*L. europaeus*) (Michell *et al.*, 2024) megfigyelthez. A havasi- és mezei nyúl genomjának összehasonlítása során a két új referencia között nagy volt a

szerveződésbeli, és kisebb a nukleotid szintű hasonlóság, a havasi nyúl kromoszóma szintű és draft genomja közöttihez képest (**4. ábra**). A draft genom scaffoldingja az OryCun2 (NCBI GenBank GCA_000003625.1) üregi nyúl referenciára alapozva történt, ami magyarázza ezt a különbséget (Marques *et al.*, 2020).

Végeredményben a teljes szekvencia hosszának 99,75%-a kromoszómához – a 23 autoszóma egyikéhez, az X, vagy az Y kromoszómához – rendelt, és a BUSCO evaluáció a keresett ortológok 95,1% (mammalia_odb10, 9226 vizsgált ortológ), illetve 93,2%-át (glires_odb10, 13798 vizsgált ortológ) tudta azonosítani. Az összeszerelt mitokondriális szekvencia 17 482 bázis hosszúságúnak adódott, ezen felül cirkuláris, és minden várt gén teljes egészében annotált rajta.



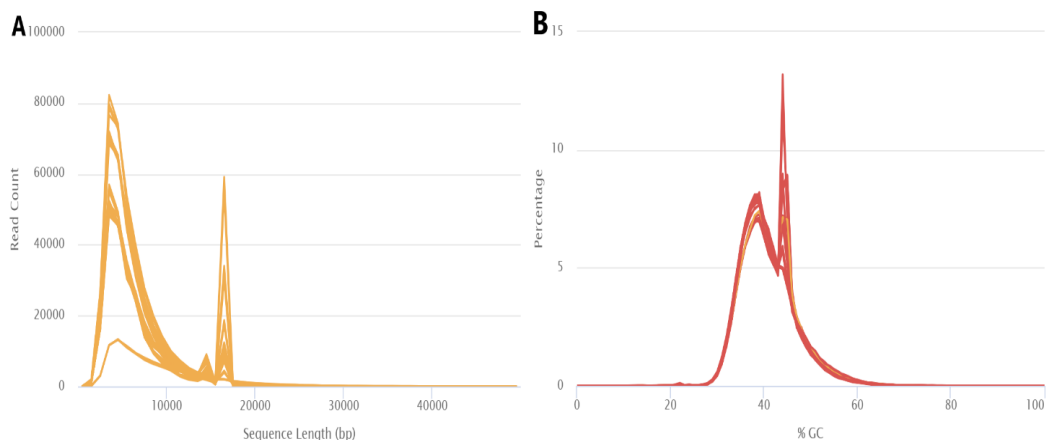
3. ábra - Azonosított telomer szekvenciák az egyes kromoszómákon. A telomer szakaszok jelenlétét piros csúcsok jelzik a kromoszóma (fekete sáv) hosszán (Location: pozíció, bázispárban megadva). A csúcsok magassága arányos az ismétlések mennyiségével. Zöld jelölés mutatja azokat a kromoszómákat, amelyek mindkét végén sikerült telomer szekvenciát azonosítani (2., 3., 4., 5., 8., 9., 14., 18. és 21 kromoszóma), narancssárga azokat, amelyeknek csak egyik végén (1., 6., 10., 11., 13., 15., 17., 19., 20. és X kromoszóma).



4. ábra - Az általam összeszerelt (*Lepus timidus timidus*, „Mountain hare (this study)”) genom összehasonlítása az A) ír havasi nyúl (*Lepus timidus hibernicus*, Irish hare) és B) mezei nyúl (*Lepus europaeus*, Brown hare) genomokkal. Az ír havasi nyúl genommal összehasonlítva jól láthatóak az üregi nyúl kromoszómaszerveződéséből adódó különbségek a pirossal kiemelt kromoszómák (Chr2, Chr1, ChrX) esetében, míg a szekvenciahasonlóság igen magas, többnyire 90% feletti. Ezzel szemben a mezei nyúl esetében alacsonyabb a szekvencia hasonlóság, és – a vártak megfelelően – kevés a kromoszóma-átrendezőzés. (Fekete et al., 2025a)

3.3. Mitokondriális DNS rekombinációja emlősökben

A mitokondriális DNS-t heteroplazmikus, MGME1 knockout sejtvonalakból nyertük ki, növelve a rekombinációnak és annak detektálásának esélyét. Az mtDNS szekvenálás előtti dúsítása ellenére a szekvenálásban nagy arányban maradtak nukleáris szekvenciák, de a GC arány és a leolvasás hosszúság eloszlásokon jól látható mitokondriális csúcsokat kaptunk (5. ábra).

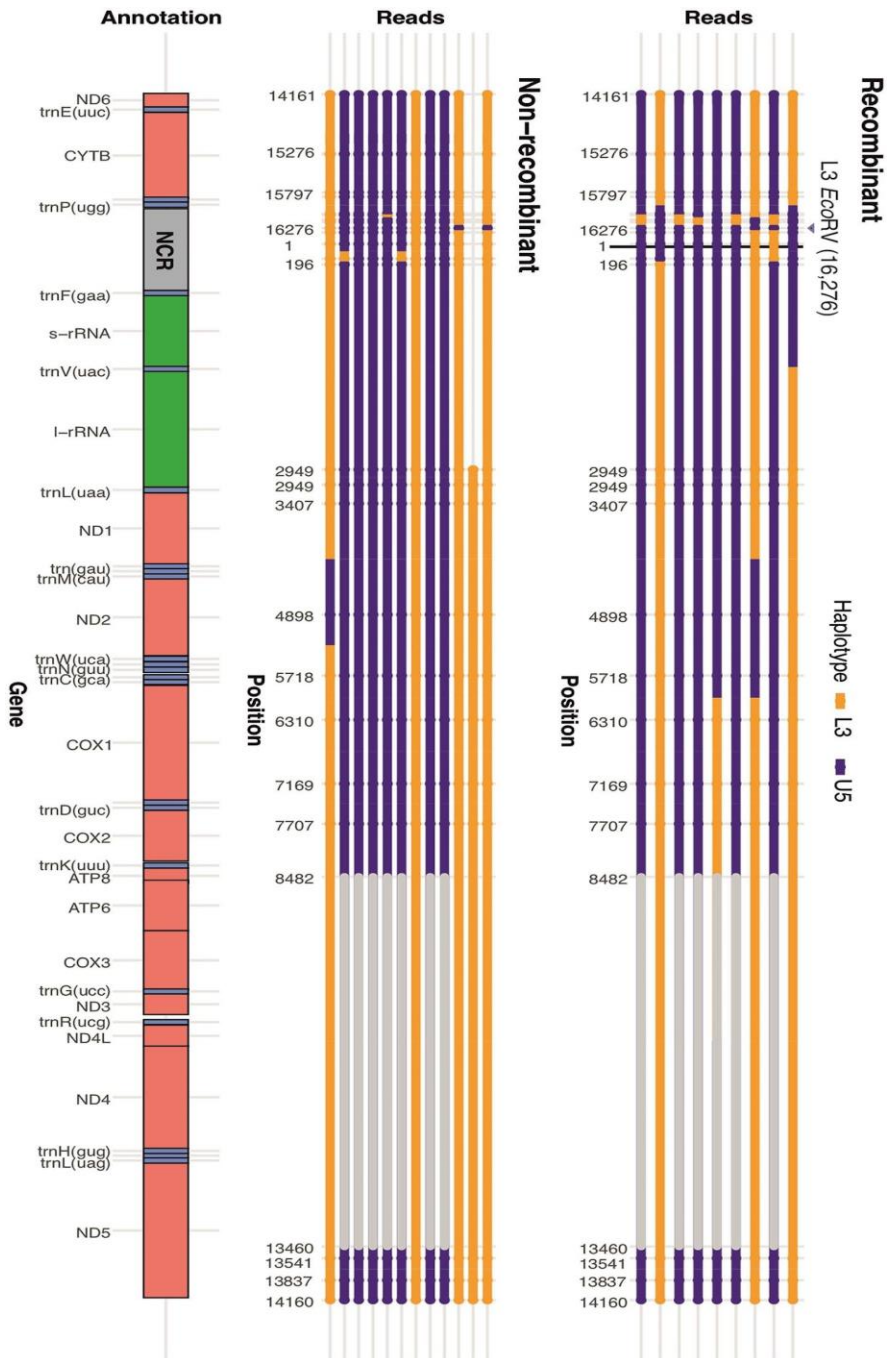


5. ábra - A) A szekvenálások leolvasás hosszúsága (Sequence length) és száma (Read Count). Jól látható a teljes mitokondriális genom hosszánál (15 és 20 kbp között) található csúcs. **B)** A leolvasások GC%-a (%GC) és az ilyen leolvasások aránya az összes leolvasás között (Percentage). A genomi leolvasásokra jellemző eloszlásból – amelyek jelen esetben kontaminációt jelentenek - erősen kiugrik a mitokondriális genomra jellemző csúcs.

Az első szűrés során a kimérikusan illeszkedett – mtDNS-re és nukleáris genomra – és a tisztán genomi szekvenciákra illeszkedő leolvasások eltávolításra kerültek. A tisztán mitokondriális szekvenciák haplotípus specifikus referenciára illesztését követően BH10 sejtvonalanban 24 SNP-t és a hosszú deléciós szakaszt, a HP202B sejtvonalanban pedig 89 SNP-t azonosítottam, amely megbízhatóan differenciáltnak adódott a haplotípusok között.

A rekombináns DNS molekulák leolvasás szintű azonosításához ezeket a variációkat vettem figyelembe, a leolvasások értékelését pedig erre a célra fejlesztett scripttel (Zenodo: <https://zenodo.org/records/10473729>) és pontozási

rendszerrel végeztem. Az alkalmazott módszerrel és beállításokkal a rekombinációs leolvasások aránya a BH10 sejtvonalonban 12%-nak (809 rekombináns / 6730 tisztán mitokondriális eredetű leolvasás), a HP202B sejtvonalonban pedig 0,8%-nak adódott (97 rekombináns / 11976 tisztán mitokondriális eredetű leolvasás). Egyes rekombináns és non-rekombináns szekvenciák vizualizációját mutatja a **6. ábra**.



6. ábra - Rekombinánsként (felül) és nem-rekombinánsként (alul) azonosított szekvenciák a BH10 sejtvonalból (Fragkoulis et al., 2024). A nem-rekombinánsként azonosított szakaszok ellenkező színnel jelölt részei olyan szakaszok, amelyek bár genotípusuk alapján az másik haplotípusba tartoznak, ezt nem tudtam kellő biztonsággal megállapítani.

4. Következtetések és javaslatok

- Bár a tiszaszentimrei, vadon élő üregi nyúl populáció nem adódott házinyúllal keresztezettnek, érdemes lehet több populációt is megfigyelni az ország területén. A magyarországi üregi nyulak populációiról kevés információ áll rendelkezésre.
- Az intergénikus, és egyéb nem kódoló, szabályozó genomi szakaszok funkciói nem jól ismertek, de a szelektált szakaszokban adódott dúsulás bizonyos transzkripciós faktor kötőhelyekre, és egyes intergénikus régiók szelektáltsága alátámasztja azok domesztikációban és szelekcióban játszott szerepét.
- A vizsgálatokat az eredmények pontosítása érdekében érdemes lehet kibővíteni, illetve megismételni egy új, nagyobb folytonosságú és teljességű referenciagenom felhasználásával.
- Az újonnan összeszerelt kromoszóma szintű havasi nyúl (*Lepus timidus*) referenciagenom jelenleg referenciaként szerepel az NCBI adatbázisában (NCBI GenBank: GCA_040893245.2).
- A havasi nyúl genom a mezei nyúl (*Lepus europaeus*) szintén csoportunk által összeszerelt referenciájával együtt jelenleg is folyamatban lévő genomikai vizsgálatok alapjául szolgál.
- Új, hosszú leolvasásokon alapuló bioinformatikai módszerrel alá tudtam támasztani az mtDNS rekombináció lehetőségét emlős sejtvonalakban. A módszer leolvasás szinten azonosítja a rekombináns molekulákat.
- Az emlős mtDNS rekombinációjának lehetségesége által nem lehet kizárni annak evolúciós jelentőségét sem. Az mtDNS rekombináció lehetőségét tehát nem vethetjük el teljesen. Különösen érdemes megvizsgálni azt olyan esetekben, ahol ezzel magyarázhatóak lehetnek egyes anomáliák.

5. Új tudományos eredmények

1. Elsőként állítottunk elő és tettünk publikussá nagy mennyiségű (42), magas lefedettségű (~35x) és jó minőségű teljesgenom-szekvenálási (WGS) adatokat több magyarországi házinyúl (*Oryctolagus cuniculus domesticus*) és üregi nyúl (*Oryctolagus cuniculus cuniculus*) populációból, amelyekről korábban nem állt rendelkezésre összehasonlítható mennyiségű genomikai információ.
2. Genomikai adatok alapján – ADMIXTURE és Treemix szoftverek, valamint D és f4 statisztikák használatával - kimutattuk, hogy a Magyarországon élő üregi nyulak (*Oryctolagus cuniculus*) és házinyulak (*Oryctolagus cuniculus domesticus*) közötti genetikai introgressziót, amely a természetes és házasított állományok közötti hibridizáció bizonyítéka.
3. Vizsgálataim során 46, a házinyúl domesztikációja során szelekció alatt álló genomi szakaszt azonosítottam, három módszer – Composite Likelihood Ratio, Fst, nukleotid diverzitás – kombinációjával.
4. A transzkripciós faktor kötőhelyek (TFBS) dúsulás elemzése során azonosítottunk olyan szabályozó szakaszokat, amelyek a házinyúl (*Oryctolagus cuniculus domesticus*) domesztikációs folyamatában valószínűleg funkcionális szerepet játszanak. Összesen 7 típusú TFBS mutatott dúsulást. Az összesen 1170 szelektáltnak adódott régióban elhelyezkedő TFBS-ből 93 génnel nem átfedő régióban található.
5. Elkészítettem a havasi nyúl (*Lepus timidus*) genomját modern bioinformatikai sztenderdek szerint, az eredményül kapott, magas minőségű genom nyilvánosan elérhető, és jelenleg ez szolgál elsődleges referenciaként az NCBI adatbázisában.

6. Általam fejlesztett, új bioinformatikai megközelítés, amelynek alkalmazásával sikerült alátámasztanom a mitokondriális DNS rekombinációjának lehetőségét, amely eddig vitatott volt emlősökben.

6. Publikációk az értekezés témakörében

1. Az értekezés témakörében megjelent tudományos cikkek

- 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., Hangan, 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/gkac040>, **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. Hivatkozások

- 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>
- Auwerwa, 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., Functamman, 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](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., Kraysberg, 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>