



Magyar Agrár- és Élettudományi Egyetem

GENOMIKAI VIZSGÁLATOK BIOINFORMATIKAI MÓDSZEREKKEL
LEPORIDAE FAJOKBAN ÉS MITOKONDRIÁLIS SZEKVENCIÁKON

Fekete Zsófia

Gödöllő

2025

A doktori iskola

megnevezése: **Állatbiotechnológiai és Állattudományi Doktori Iskola**

tudományága: Agrártudományok, Állattenyésztési tudományok

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

RÖVIDÍTÉSEK JEGYZÉKE	5
1. BEVEZETÉS	8
2. CÉLKITŰZÉSEK	10
2.1. Populációgenomikai vizsgálatok üregi nyúlban.....	10
2.2. Havasi nyúl (<i>Lepus timidus</i>) referenciagenom összeszerelése.....	10
2.3. Lineáris DNS-vezérelt rekombináció vizsgálata mitokondriális szekvenciákon	10
3. IRODALMI ÁTTEKINTÉS.....	11
3.1.Szekvenálási módszerek a populációgenomikában és -genetikában.....	11
3.1.1. A szekvenálás elterjedése a genetikai vizsgálatokban	11
3.1.2. NGS: A következő generáció.....	11
3.1.3. A harmadik generáció: egymolekula-szekvenálások.....	12
3.1.4. Szekvenálási technológiák a populációgenomikában.....	12
3.1.5. Szekvenálási technológiák a genom összeszerelésben	13
3.2. A vizsgált fajok – A Leporidae család <i>Oryctolagus</i> és <i>Lepus</i> nemzetsége	14
3.3. Hibridizáció és szelekció	15
3.3.1. A domesztikáció hatásai	15
3.3.2. A házi és vad populációk közötti hibridizáció lehetősége és potenciális hatásai	16
3.3.3. Hibridizáció és introgresszió a természetben.....	16
3.3.4. A hibridizáció genomi lenyomatai.....	17
3.3.5. A hibridizáció detektálása.....	18
3.3.6. A szelekció genomi lenyomatai	19
3.3.7. A szelekció detektálása	20
3.4. A rekombináció evolúciós szerepe és azonosítási módszerei	21
3.4.1. A rekombináció mechanizmusai.....	22
3.4.2. Rekombináció azonosítása a genomon	23
3.4.3. Mitokondriális rekombináció.....	23
3.5. A genomikai vizsgálatok háttérének megteremtése - A referenciagenomok	24
3.5.1. Biológiai kihívások a referenciagenomok összeszerelésében.....	25
3.5.2. A referenciagenom összeszerelés folyamata	26
3.5.3. Referenciák minőségi értékelése.....	27
3.5.4. A genomösszeszerelés és a lineáris referenciagenomok kihívásai és limitációi.....	28
3.5.5. A referencia genomok hatása az analízisre	29
4. ANYAG ÉS MÓDSZER	31
4.1. Teljes genom szekvenálás alapú populációgenetikai vizsgálatok vadon élő és házinyúl fajtákban.....	31
4.1.1. Mintagyűjtés és DNS izolálás.....	31
4.1.2. Szekvenálás.....	31

4.1.3. Publikus adatok gyűjtése.....	31
4.1.4. Minőségi kontroll és genetikai variációk azonosítása	32
4.1.5. Populáció-struktúra és beltenyésztettség	33
4.1.6. Szelektált genomi régiók detektálása és jellemzése házinyúl fajtákban.....	34
4.2. Kromoszóma szintű <i>Lepus timidus</i> referenciagenom összeszerelése.....	34
4.2.1. Mintagyűjtés és fajazonosság megerősítése.....	34
4.2.2. DNS izolálás és szekvenálás	35
4.2.3. Minőségi kontroll.....	35
4.2.4. Genom összeszerelés	36
4.2.5. Telomer- és repetitív szakaszok annotációja	36
4.2.6. Kézi kuráció	36
4.2.7. Mitokondriális szekvencia összeszerelése	37
4.2.8. Összehasonlítás elérhető <i>Lepus</i> genom szekvenciákkal	37
4.3. Lineáris DNS-vezérelt mitokondriális rekombináció kimutatása emlős sejtekben ...	37
4.3.1. Felhasznált sejtvonalak	37
4.3.2. DNS izolálás és szekvenálás	37
4.3.3. Minőségi ellenőrzés és tényleges mitokondriális szekvenciák szelektálása.....	38
4.3.4. Mitokondriális haplotípusok közötti megbízható különbségek azonosítása.....	39
4.3.5. Rekombináns szekvenciák azonosítása – az alkalmazott script és működése.....	40
5. EREDMÉNYEK ÉS DISZKUSSZIÓ	41
5.1. Házi- és üregi nyúl hibridizációja, valamint szelekció a nyúl házasítása során	41
5.1.1. Szekvenálási eredmények és minőségi szűrés	41
5.1.2. Populáció-struktúra	42
5.1.3. Hibridizáció vad- és házinyulak között.....	46
5.1.4. A beltenyésztettség mértéke	48
5.1.5. Szelekció által érintett régiók.....	50
5.1.6. Szelektált gének és variációk házinyúl fajtákban	53
5.2. Új havasi nyúl (<i>Lepus timidus</i>) referenciagenom összeszerelése	57
5.2.1. Szekvenálási eredmények és genom összeszerelés.....	57
5.2.2. Az összeszerelt genom.....	59
5.2.3. A genom minőségi ellenőrzése	62
5.2.4. A mitokondriális szekvencia.....	63
5.2.5. Az adatok elérhetősége	64
5.2.6. Az új <i>Lepus timidus</i> referencia összevetése korábban összeszerelt <i>Lepus</i> genomokkal	64
5.3. Mitokondriális rekombináció előfordulásának megerősítése emlős sejtekben	66
5.3.1. Sejtvonalak és szekvenálás	66
5.3.2. A szekvenálás minőségi ellenőrzése és nem-mitokondriális DNS szűrése	67
5.3.3. Haplotípus különbségek - megbízható változatok	68
5.3.4. A rekombináció kimutatása mitokondriális DNS szekvenciákon	69
5.3.5. A laboratóriumi eredmények sikeres igazolása bioinformatikai módszerrel.....	71
5.3.6. A rekombináció jelenléte a mitokondriális genomban	71
6. KÖVETKEZTETÉSEK ÉS JAVASLATOK	73
6.1. Domesztikáció és szelekció üregi nyúlban	73
6.1.1. Házi- és vad üregi nyúl állományok hibridizációja	73

6.1.2. Poligenikus domesztikáció és szabályozó szakaszok szerepe	73
6.2. A havasi nyúl (<i>Lepus timidus</i>) referenciagenom	73
6.3. Mitokondriális rekombináció alátámasztása emlős sejtvonalakban	74
7. ÚJ TUDOMÁNYOS EREDMÉNYEK.....	75
8. ÖSSZEFOGLALÁS	76
8.1. Bevezetés.....	76
8.2. Anyag és módszer	76
8.2.1. Teljes genom szekvenálás alapú populációgenomikai vizsgálatok üregi- és házinyúlban	76
8.2.2. <i>Lepus timidus</i> referenciagenom összeszerelése.....	77
8.2.3. Mitokondriális DNS rekombinációjának igazolása emlős sejtekben.....	77
8.3. Eredmények és diszkusszió	78
8.3.1. Hibridizáció és szelekció az üregi- és házinyúl között	78
8.3.2. Havasi nyúl (<i>Lepus timidus</i>) referenciagenom	78
8.3.3. Mitokondriális DNS rekombinációja emlősökben.....	79
8.4. Következtetések és javaslatok.....	79
9. SUMMARY	81
9.1. Introduction	81
9.2. Materials and methods.....	81
9.2.1. Whole genome sequencing based population genetic analysis of wild and domestic rabbits.....	81
9.2.2. New, high-quality reference genome for the mountain hare (<i>Lepus timidus</i>).....	82
9.2.3. Detecting the recombination of mitochondrial DNA in mammalian cell cultures	82
9.3. Results and discussion	83
9.3.1. Hybridization and selection between the wild and domestic European rabbit	83
9.3.2. A high-quality reference genome of the mountain hare (<i>Lepus timidus</i>).....	83
9.3.3. Mitochondrial DNA recombination in mammalian cells.....	84
9.4. Conclusions and suggestions.....	84
MELLÉKLETEK.....	86
M1. IRODALOMJEGYZÉK	86
M2. ORYCTOLAGUS CUNICULUS MINTÁK JEGYZÉKE.....	113
M3. ERŐSEN DIFFERENCIÁLT VARIÁCIÓK A SZELEKTÁLT RÉGIÓKBAN	117
M4. TRANSZKRIPCIÓS FAKTOR KÖTŐHELYEK A SZELEKTÁLT RÉGIÓKBAN	138
M5. MITOKONDRIÁLIS REKOMBINÁCIÓ – SCRIPTEK.....	152
1. mtrec_detector.py – main script.....	152
2. ioprocessing.py – Input/output format processing functions	154
3. seqhandlers.py – Sequence and data processing and calculations.....	155
4. Kiegészítő scriptek.....	157

4.1. select_pure_mt.py – Tisztán mitokondriális readek kiválasztása az illesztésből.....	157
4.2. Ábrák generálásához használt R scriptek	158
M6. D ÉS F4 TESZTEK EREDMÉNYEI – ÜREGI- ÉS HÁZINYÚL HIBRIDIZÁCIÓ	162
.....	
A) Tesztek a wildHU_b (Budapesti Állat- és Növénykert) üregi nyúl populációra	162
B) Tesztek a wildHU_a (Tiszaszentimre) üregi nyúl populációra.....	162
C) Összevont tesztek az összes magyarországi üregi nyúl mintával	163
M7. LEPUS TIMIDUS GENOM-ÖSSZESZERELÉS – STATISZTIKÁK	164
1. Az összeszerelt genom számszerű jellemzői (Fekete <i>et al.</i>, 2025a)	164
2. BUSCO eredmények (Fekete <i>et al.</i>, 2025a).....	165

RÖVIDÍTÉSEK JEGYZÉKE

ATP2B1	– ATPase Plasma Membrane Ca ²⁺ Transporting 1, gén
ATRNL1	– Attractin-like 1, gén
BCL6	– B-Cell lymphoma 6, gén
BIR	– Break-induced repair, kettős-száltörés indukált replikáció
Bp	– bázispár
BQSR	– Base Quality Score Recalibrator
CEP85	– Centrosomal Protein 85, gén
CFLAR	– CASP8 And FADD Like Apoptosis Regulator, gén
CLR	– Composite Likelihood Ratio
ERGA	– European Reference Genome Atlas
FDR	– False discovery rate
FHIP1A	– FHF Complex Subunit HOOK Interacting Protein 1A, gén
FISH	– Fluoreszcens in-situ hibridizáció
FR	– French, franciaországi üregi nyúl populációk jelölése
GATK	– Genome Analysis Toolkit
GO	– Gene ontology
Hi-C	– High throughput chromosome conformation capture
HU	– Hungarian, magyarországi üregi nyúl populációk
HR	– Homológ rekombináció
IB	– Iberian, Ibériai-félszigeti üregi nyúl populációk
Indel	– inszerciók és/vagy deléciók együttesen
IUCN	– International Union for Conservation of Nature
JW	– Japanese White, japán fehér házinyúl fajta
LCORL	– Ligand Dependent Nuclear Receptor Corepressor Like, gén
LD	– Linkage disequilibrium, kapcsoltsági egyensúlyhiány
lncRNS	– hosszú nem kódoló RNS
LR	– Long repeat
LT1	– <i>Lepus timidus</i> 1, immortalizált fibroblaszt sejtvonal

MAP3K21	– Mitogen-Activated Protein Kinase Kinase Kinase 21, gén
MAN1C1	– Mannosidase Alpha Class 1C Member 1, gén
MGME1	– Mitochondrial genome maintenance exonuclease 1, gén
MSL2	– MSL (male-specific lethal homolog) Complex Subunit 2, gén
mtDNS	– mitokondriális DNS
NCBI	– National Center for Biotechnology Information
NGS	– Next Generation Sequencing, Következő Generációs Szekvenálás
NHEJ	– Non-homologous End Joining, Nem-homológ végek összekapcsolása
Nldwarf	– Netherlands Dwarf, holland törpe házinyúl fajta
NZB	– New Zealand Black, egér fajta
NZW	– New Zealand White, új-zélandi fehér házinyúl fajta
ONT	– Oxford Nanopore Technologies
PacBio	– Pacific Biosciences, illetve az általuk fejlesztett és forgalmazott szekvenálási technológiák
PCNX2	– Pecanex 2, gén
POC1B	– POC1 Centriolar Protein B, gén
POLG	– DNA polymerase gamma, catalytic subunit
POLR1B	– DNA-directed RNA polymerase I subunit RPA2, gén
PPP2R3A	– Protein Phosphatase 2 Regulatory Subunit B"Alpha, gén
RADseq	– Restriction site associated DNA sequencing
ROH	– Run of homozygosity – Homozigóta genomi szakasz
SBS	– Sequencing by synthesis – Szekvenálás szintézis által
SCN1A	– Sodium channel protein type 1 subunit alpha, gén
SCN7A	– Sodium channel protein type 7 subunit alpha, gén
SCN9A	– Sodium channel protein type 9 subunit alpha, gén
SDSA	– Synthesis-dependent Strand Annealing
SH3BGRL3	– SH3 Domain Binding Glutamate Rich Protein Like 3, gén
SIPA1L2	– Signal-Induced Proliferation-Associated 1-Like Protein 2, gén
SMRT	– Single Molecule Real-time Sequencing
SNP	– Single Nucleotide Polymorphism – egynukleotidos polimorfizmus
SRA	– Sequence Read Archive
TEM	– Transmission electron microscopy – Transzmissziós elektron-mikroszkópia

TGS	– Third Generation Sequencing, Harmadik Generációs Szekvenálás
TFBS	– Transcription factor binding site – Transzkripciós faktor kötőhely
UBXN11	– UBX Domain Protein 11, gén
VGP	– Vertebrate Genome Project
VQSR	– Variant Quality Score Recalibration
WGS	– Whole Genome Sequencing – Teljes genom szekvenálás
wildFR	– vadon élő, Franciaország területéről származó üregi nyúl populáció
wildHU_a	– vadon élő, Magyarország (Tiszaszentimre) területéről származó üregi nyúl populáció
wildHU_b	– Magyarországról (Budapesti Állat- és Növénykert) származó üregi nyúl populáció
wildIB	– vadon élő, az Ibériai-félsziget területéről származó üregi nyúl populáció

1. BEVEZETÉS

A természetes és mesterséges szelekciós folyamatok, a fajok és fajták hibridizációja, az introgresszió vagy a rekombináció jelenléte mind vizsgálható modern szekvenálási technológiákkal. A rövid- és hosszú leolvasások, az amplifikációs, célzott vagy a teljes genom szekvenálására alapuló vizsgálatok és ezek kombinációja számos felhasználási lehetőséget rejt magában. Dolgozatom három témájából kettőben elsősorban *Leporidae* fajok populációgenomikai vizsgálatával és az ehhez rendelkezésre álló referenciákkal foglalkozom, míg harmadik témaként a rekombináció lehetőségének felderítésével foglalkozom emlős mitokondriális genomokban.

Az üregi nyúl (*Oryctolagus cuniculus*) és házi nyúl formáján, a házi nyúl végzett populációgenomikai vizsgálataim céljára felhasznált Illumina szekvenálások magas lefedettsége és jó minősége ellenére többször éreztem a referenciagenom hiányosságait. Az OryCun2.0 referenciagenom egy Thorbecke inbred vonalból származó nőstény házi nyúl egyedből készült, shotgun szekvenálási technikával (NCBI RefSeq assembly GCF_000003625.3, GenBank accession GCA_000003625.1). A rövid leolvasások használata töredezett, a kromoszómák mellett sok rövid, lokalizálatlan szekvenciát tartalmazó referenciát eredményezett, amely több szempontból is megnehezíti az elemzéseket. Csak 2024 augusztusában jelent meg az új mOryCun1.1 referencia (NCBI RefSeq assembly GCF_964237555.1, GenBank accession GCA_964237555.1), amely teljességében és folytonosságában is a (Rhie *et al.*, 2020a) által lefektetett modern szenderdeknek megfelelő.

Hiányosságai ellenére az üregi nyúl nagyon sokáig az egyetlen *Leporidae* faj volt, amelyhez egyáltalán rendelkezésre állt kromoszóma szinten összeszerelt referenciagenom, így számos kutatásban került felhasználásra más *Leporidae* fajok esetében is, úgynevezett „pseudo-referenciák” alapjaként. A *Oryctolagus* és *Lepus* nemzetségek genomszerveződése azonban eltér annyira, hogy bár az üregi nyúl genom jobb híján használható referenciaként, ez messze legyen az ideális megoldástól. A legszembetűnőbb különbség a nemzetségek eltérő kariotípusa: a *Lepus* fajok azonos számú kromoszómával rendelkeznek ($2n=48$), az üregi nyúlé azonban ettől eltérő ($2n=44$).

Ma már több tudományos konzorcium létezik, amely a modern szenderdeknek megfelelő referencia genomok összeszerelésével és ezek további felhasználásával foglalkozik, például az ERGA (European Reference Genome Atlas), amely európai fajokra fókuszál, de ennél szűkebb fókusszal operáló konzorciumok is léteznek, mint a VGP (Vertebrate Genome Project), ASG (Aquatic Symbiotic Genomes consortium), vagy a Project Psyche, amely Lepidoptera genomokra fókuszál. E csoportok közül kutatócsoportjaink többel is együttműködnek, megteremtve az általam is végzett genom összeszerelési munka, a *Lepus* nemzetségbe tartozó első két, kromoszóma szinten összeszerelt referenciagenom – havasi nyúl (*Lepus timidus*) (Fekete *et al.*, 2025a) és mezei nyúl (*Lepus europaeus*) (Michell *et al.*, 2024) – háttérét is.

A mezei nyúl és havasi nyúl közti hibridizáció jelenléte jól ismert, de hatásai jelenleg is kutatás tárgyát képezik. A Finnország területén létező kiterjedt hibrid zóna ideális terület ennek megfigyelésére. Az ezt célzó kutatás első lépése volt a dolgozatomban tárgyalt havasi nyúl referencia létrehozása is. Az új referenciagenomok lehetőséget adnak közeli rokon fajok evolúciós vizsgálatára, és a későbbiekben a két hibridizálódó faj kiterjedt és pontos vizsgálatára.

A hibridizáció és szelekció vizsgálata az üregi nyúl esetében fontos téma, ahol a házi nyúl nemesítés és az esetleges vad üregi nyulakkal való hibridizáció kérdéseire keresünk választ. A

magyarországi üregi nyulak helyzetéről kevés információ áll rendelkezésre, de felmerült a hibridizáció lehetősége a házi- és vad populációk között, amely befolyásolhatja a vadon élő populációk fitnessét. Ezen kívül, bár számos tanulmány vizsgálja az egyes házinyúl fajtákban szelekció alatt álló géneket, és tudjuk, hogy a populáció a házasítás folyamán és azt követően is ki lett téve palacknyak (bottleneck) hatásnak, a domesztikáció általános hatása és az üregi- és vad populációk közötti differenciáció kevésbé vizsgált az egyes fajtákénál.

Dolgozatom harmadik témája a lineáris DNS vezérelt mitokondriális rekombináció lehetősége emlős mitokondriális genomokban. A mitokondriális rekombináció előfordulása, bár számos fajban ismert és elfogadott, akár evolúciós időskálán, illetve jelentőséggel is (Ladoukakis és Zouros, 2001; Ladoukakis *et al.*, 2011) emlősökben ennek lehetősége az mtDNS maternális öröklődése miatt erősen korlátozott, és ebből adódóan vitatott. Korábban több esetben megfigyeltek mitokondriális rekombinációt humán sejtekben is (Kraytsberg *et al.*, 2004; Zsurka *et al.*, 2005; 2007).

Csoportunk laboratóriumi körülmények között sikeresen alátámasztotta a mitokondriális rekombináció lehetőségét heteroplazmikus sejtvonalakban. Ezekben mikroszkópos, molekuláris biológiai, és bioinformatikai módszerekkel is ki tudtuk mutatni rekombináns mitokondriális szekvenciák jelenlétét. Ez utóbbihoz a fentebb említett, korábbi kutatások idejében még nem elérhető, nagy pontosságú, hosszú leolvasásos szekvenálási technikát alkalmaztunk, kiegészítve a rekombináció leolvasás szintű azonosítására alkalmas, új scriptekkel.

A mitokondriális rekombináció előfordulásának alacsony valószínűsége miatt biológiai hatása nem ismert és nem vizsgált, bár a szelekciós folyamatok fontos tényezője, megteremtője a rekombináció. A mezei- és a havasi nyúl hibrid zónájában gyakran megfigyelt a két faj közötti hibridizáció eredményeként a mitokondriális introgresszió, túlnyomó részben a havasi nyúl felől a mezei nyúl irányába. Az introgresszió számos esete (Levänen *et al.*, 2018a) mellett mitokondriális genom LR (long repeat) régiójára heteroplazmikus egyedeket is megfigyeltek (Tapanainen *et al.*, 2024) a hibrid zónában. Kutatócsoportunk témái közé tartozik a mitokondrium szerepének vizsgálata az adaptációban – a rekombináció lehetősége pedig további érdekes kérdéseket vethet fel.

2. CÉLKITŰZÉSEK

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

2.1. 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 populációról, azok eredetéről, illetve pillanatnyi állapotáról. Célul tűztem ki a vadon élő üregi nyulak helyzetének vizsgálatát Magyarországon, beleértve az esetleges hibridizációt házi nyulakkal és a beltenyésztettség mértékét.

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

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

Célul tűztem ki egy, eddig referenciagenommal nem rendelkező Leporid faj (havasi nyúl, *Lepus timidus*) genomjának modern sztenderdeknek megfelelő, referencia minőségű összeszerelését a pillanatnyilag elfogadott legjobb gyakorlatot követve, és az összeszerelt genom összevetését egy egy hibridizálódó közeli fajnak, a mezei nyúlnak (*Lepus europaeus*) genomjával.

2.3. Lineáris DNS-vezérelt rekombináció vizsgálata mitokondriális szekvenciákon

Célul tűztem ki ezen felül egy bioinformatikai módszer fejlesztését rekombináció detektálására hosszú read (PacBio) szekvenálásokból, 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. Az általam kifejlesztett módszer célja a rekombináció előfordulásának bizonyítása emlős sejtekben, mitokondriális szekvenciákon, és ez által laboratóriumi eredmények bioinformatikai hitelesítése.

3. IRODALMI ÁTTEKINTÉS

3.1. Szekvenálási módszerek a populációgenomikában és -genetikában

3.1.1. A szekvenálás elterjedése a genetikai vizsgálatokban

A szekvenálási módszerek hosszú utat tettek meg Sanger és Coulson „plusz-mínusz” módszerének kifejlesztése óta (Sanger és Coulson, 1975). Az eljárás továbbfejlesztését, majd részleges automatizálását (Sanger *et al.*, 1977; Smith *et al.*, 1986) követően hamar megjelentek az első, kereskedelmi forgalmú alkalmazások és szekvenáló készülékek is.

A shotgun szekvenálás bakteriális vektorokat használ a DNS véletlenszerűen fragmentált darabjainak klónozására, így a szekvenciákat párhuzamosan lehetséges szekvenálni, és az átfedések alapján összeilleszteni a leolvasásokat (Messing *et al.*, 1981), ez jelentős áteresztőképesség növekedéshez vezetett. Hamarosan olyan, további mérföldkövek segítettek a DNS szekvenálás rohamos fejlődését, mint a PCR (Mullis *et al.*, 1986) a korábbinál hatékonyabb fluoreszcens jelölés (Metzker *et al.*, 1996), a mágneses gyöngyök használatával történő DNS tisztítás (DeAngelis *et al.*, 1995) és a kapilláris elektroforézis a fluoreszcens jelek azonosítására (Zhang *et al.*, 1995).

A 1980-as évektől kezdve egyre több teljes genom szekvenálási projekt indult. A növekedés mértékét jól illusztrálja a GenBank szekvencia-adatbázis 1982-es alapítását követő rohamos növekedése (<https://www.ncbi.nlm.nih.gov/genbank/statistics/>, 2025.05.09).

3.1.2. NGS: A következő generáció

A „következő generációs” szekvenálási módszerek (Next Generation Sequencing, NGS) jellemzője az áteresztőképesség rohamos növekedése az új technológiáknak köszönhetően. Közös vonásuk a DNS fragmentációját és amplifikációját követő masszív parallel szekvenálás (Head *et al.*, 2014).

A pirosekvencálás (Margulies *et al.*, 2005) során a DNS-szintézis során felszabaduló pirofoszfátok által kibocsátott szignál alapján történik a detektálás, a SOLiD (Brenner *et al.*, 2000) specifikus DNS-ligázok és fluoreszcens oligonukleotidok használatára alapszik, az IonTorrent (Rothberg *et al.*, 2011) pedig protonok detektálására félvezetők által. A SOLiD és az IonTorrent problémákba ütközött palindrom és homopolimer szekvenciák esetében (Huang *et al.*, 2012; Loman *et al.*, 2012), amelyek megoldása hiányában a technológiák idővel eltűntek a piacról.

A verseny nyertese végül a „sequencing by synthesis” (SBS) eljárás lett. Az eljárás során azonos DNS fragmensek felületen rögzített klaszterei adják a templát könyvtárat. Az amplifikáció felerősíti a kapott jelet, megkönnyítve ezzel a bázisok hívását, amely nagyobb pontosságot eredményez, a detektáláshoz szükséges optikából adódó költség jelentős csökkenése mellett, továbbá a reverzibilis terminátorok használatával folytonos szintézis lehetséges (Adessi, 2000). Az Illumina mai készülékei már 8000 Gbp adatot képesek generálni egy futás (17-48 óra) alatt (<https://www.illumina.com/systems/sequencing-platforms.html>, 2025.05.09).

Az NGS elterjedése számos új alkalmazást is lehetővé tett, úgymint a nagy áteresztőképességű RNS szekvenálás (Nagalakshmi *et al.*, 2008) a teljes exom szekvenálás (Ng *et al.*, 2009), a Chip-Seq (chromatin-immunoprecipitációt követő szekvenálás) (Johnson *et al.*, 2007), epigenetikai

jelek (Lister *et al.*, 2008) vagy a kromatin konformációjának és az egyes genomi szakaszok fizikai közelségének vizsgálata (Lieberman-Aiden *et al.*, 2009).

3.1.3. A harmadik generáció: egymolekula-szekvenálások

A harmadik, vagy legújabb generációs szekvenálások alatt általában az amplifikáció nélküli, egyedi DNS molekulák szekvenálására képes technológiákat értünk.

Az első egymolekulás szekvenálási technológia SBS (sequencing by synthesis) technológián alapult (Braslavsky *et al.*, 2003; Harris *et al.*, 2008), ám relatív lassúsága, magas költsége, és az eredményként kapott aránylag rövid szekvenciák miatt nem tudott elterjedni.

Mára a terület két fő versenyzője a Pacific Biosciences (PacBio) (Rhoads és Au, 2015) és az Oxford Nanopore Technologies (ONT) (Jain *et al.*, 2016), az általuk piacra dobott technológiákkal.

A PacBio egymolekulás valós idejű szekvenálása (Single Molecule Real-Time sequencing, SMRT) általában 0-20 kbp medián hosszúságú leolvasásokat generál, melyek akár 50 kbp hosszúságúak is lehetnek. A könyvtárkészítés során hairpin-adaptereket ligálnak a DNS fragmens két végéhez. Az ilyen módon cirkularizált molekulákon folyamatosan több leolvasást lehet végezni, amelyből aztán a konszenzus szekvencia (circular consensus sequence, CCS) leolvasás kialakul. A módszer kezdeti pontatlansága (kb 1 hiba / 10 nukleotid) ellenére a szisztemikus hiba hiánya, a többszörös leolvasás által lehetséges pontos konszenzus, valamint a jelfeldolgozás és a szekvenálási kémia fejlődésének köszönhetően mára az Illumina leolvasásokéhoz hasonló pontosságot ér el a technológia (Rhoads és Au, 2015; Wenger, A. M. *et al.*, 2019).

A Nanopore technológia genetikailag módosított bakteriális nanopórusok használatára alapul. A nanopórusok egy mesterséges kettős lipidmembránon helyezkednek el, a DNS szálak ezeken a pórusokon keresztül mozognak. A molekula áthaladása megzavarja a póruson áthaladó elektromos áramlást, amelynek mértéke az egyedi nukleotidokra jellemző, és amelyet egy félvezető szenzor regisztrál (Deamer *et al.*, 2016). Az átlagos leolvasás hosszúság hasonló a PacBio-hoz, de a leolvasás hosszúságát elsősorban az izolált DNS hossza befolyásolja. A leghosszabb szekvenált molekula elérte 2,3 Mbp-t (Payne *et al.*, 2019), emellett a Nanopore technológia potenciálisan képes olyan repetitív régiók szekvenálására, amelyekre más módszerek nem. A módszer pontossága is jelentős fejlődésen ment át az elmúlt pár évben (Sereika *et al.*, 2022).

3.1.4. Szekvenálási technológiák a populációgenomikában

A legegyszerűbb, legkevesebb anyagi és technológiai erőforrást igénylő módszerek közé tartozik egyedileg kiválasztott lókuszok célzott szekvenálása. Egy ilyen típusú vizsgálathoz nem feltétlenül szükséges referenciagenom használata sem: a szekvenált lókuszok között többszörös illesztéssel könnyen azonosíthatóak a különbségek. Bár egyszerű és költséghatékony, a genomról kapott információ ebben az esetben erősen limitált.

Ennél jóval nagyobb mennyiségű információt szolgáltatnak a restrikciós technikákon alapuló vizsgálatok. A RAD (Restriction site Associated DNA) szekvenálás (Peterson *et al.*, 2012) és változatai, valamint a „genotyping-by-sequencing” (GBS) (Wallace és Mitchell, 2017) egyszerre számos genomi régió vizsgálatára alkalmas, de a genom lefedettsége a restrikciós emésztés természetéből adódóan továbbra sem egyenletes, vagy teljes. Költséghatékonyasága miatt nagy mennyiségű, akár több száz egyed vizsgálatát teszi lehetővé. A restrikciós szekvenálás eredményei felhasználhatók ugyan referenciagenom használata nélkül – a Stacks programcsomag például képes RADseq adatokból lokális *de novo* összeszerelés alapján variációk hívására (Catchen *et al.*,

2013) –, a megfelelő referenciagenom megkönnyíti az analízist és javítja az eredmények megbízhatóságát (Bohling, 2020).

A restriktív technikákkal szemben a teljes genom szekvenálás (whole genome sequencing, WGS) (Lu *et al.*, 2025) randomizált, nagyobb volumenű, és emiatt egyenletesebb lefedettsége elérhető. A genom szinte teljes egészéről jó minőségű, nagy megbízhatóságú információt nyerhetünk, sok minta esetében azonban a WGS igen költségessé válik. Egyes minták poolozása ugyan lehetővé teszi a több egyedből származó szekvencia adatok kinyerését, ez azonban akadályoz olyan vizsgálatokat, amelyekhez egyedi genotipizálás szükséges. Az adatok feldolgozásához és kiértékeléséhez elengedhetetlen a referenciagenom.

A transzkriptomikai, RNS-szekvenálással végzett vizsgálatok limitációja, hogy az RNS expresszió szövetspecifikussága és más tényezők általi befolyásoltsága miatt (pl.: cirkadián ritmus, stressz körülmények) miatt csak a mintavétel helyén és idején jelen lévő RNS-ekről jutunk információhoz. Ráadásul a detektált expressziós szintet a szekvenálás egyenetlensége, az amplifikációs lépésből adódó torzítás is befolyásolja. Emiatt a transzkriptomikai módszerek inkább kiegészítő módszerként alkalmasak evolúciós- és populációgenomikai vizsgálatokban.

3.1.5. Szekvenálási technológiák a genom összeszerelésben

Az első nagy méretű emlős genomok szekvenálása összeszerelése shotgun szekvenáláson alapult. A Humán Genom Projektet követően pár éven belül a legtöbb modell organizmushoz is elkészültek referenciák (Giani *et al.*, 2020).

A referenciagenomok párhuzamosan fejlődtek a szekvenálási technikákkal: a „következő generáció” már nagy áteresztőképességű technikákkal készült. Ezek hátránya, hogy bár rövid, például bakteriális genomokon könnyen elérhető nagy lefedettség, és a leolvasások kellő mértékű átfedése a hatékony összeszereléshez, de minél hosszabb a genom, ezt annál nehezebb és költségesebb elérni. Különösen nagy problémát jelentettek a rövid leolvasásoknál a repetitív genomi szakaszok (Biscotti *et al.*, 2015). Ez utóbbi problémán a harmadik generációs szekvenálások megjelenése ugyan javított, de teljesen kiküszöbölni a mai napig sem sikerült.

Az egymolekula-szekvenálásos, hosszú leolvasásokat eredményező és a könyvtárkészítés során amplifikációs lépés nélküli technológiákkal együtt megjelentek a hibrid technikák, amelyek rövid- és hosszú leolvasások kombinációjával dolgoztak (Minoche *et al.*, 2011; Zimin *et al.*, 2017). Mára a rövid leolvasások támogató szerepbe szorultak vissza, de nagy pontosságuknak, illetve a rájuk épülő, genom konformációt mutató technológiáknak, mint amilyen a Hi-C (high throughput chromosome conformation capture) (Lieberman-Aiden *et al.*, 2009; Belaghal *et al.*, 2017) köszönhetően nem szorultak ki teljesen. A Hi-C módszer során a szekvenálást megelőzően egymáshoz közel lévő DNS szakaszok között keresztkötéseket hoznak létre. Ezt követi a fragmentálási lépés, majd az eltört végek újbóli ligálása, amellyel kimérikus szekvenciák jönnek létre. Az egyes termékek abundanciája képet ad az eredeti DNS szakaszok fizikai közelségéről.

A referenciagenomok összeszerelésénél ma három fő irányelv érvényesül, ezekből kettő közvetlenül az alkalmazott szekvenálási és térképezési módszerekre vonatkozik. Ezek: lehetőség szerint a kiindulási adatok egy része hosszú-leolvasásos szekvenálásból származzon; a Hi-C és az optikai térképezés közül legalább egy módszer kerüljön felhasználásra a scaffoldinghoz; az adatokat és az elkészült genomot pedig vessék alá gondos evaluációnak (Whibley *et al.*, 2021).

3.2. A vizsgált fajok – A Leporidae család *Oryctolagus* és *Lepus* nemzetsége

Munkám során főleg Leporidae családba tartozó fajokkal foglalkoztam, ezek közül is az *Oryctolagus* és *Lepus* nemzetségbe tartozóakkal.

Az *Oryctolagus* nemzetség ma élő egyetlen tagja az *Oryctolagus cuniculus*, az üregi nyúl. Őshazájában, az Ibériai-félszigeten két alfaja, az *O. c. cuniculus* és az *O. c. algirus* is megtalálható. Az üregi nyúl érdekessége, hogy egyszerre létezik háziásított, és vad formában is, amely remek lehetőséget kínál evolúciós és szelekciós vizsgálatokra.

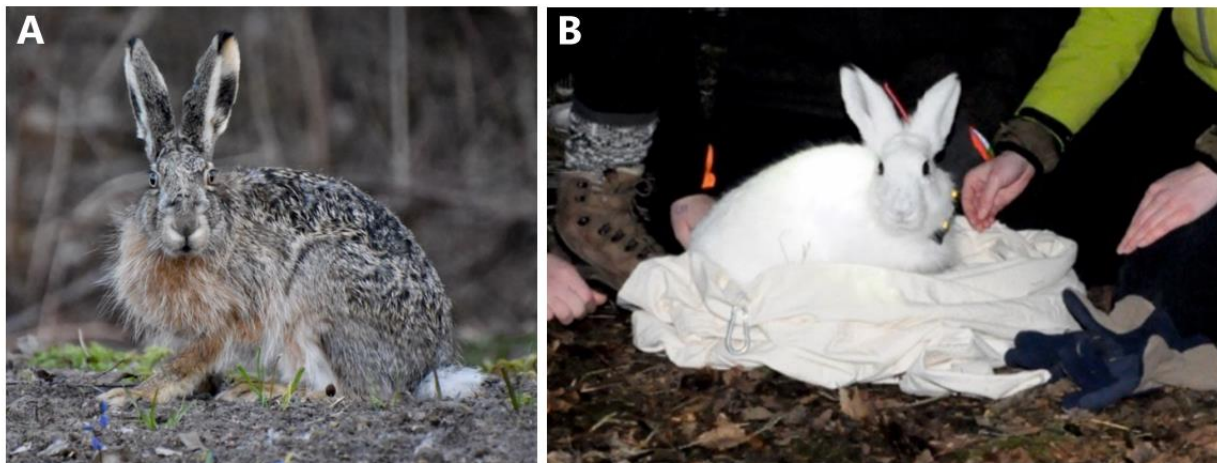
Az üregi nyúl sokoldalú faj. Vad formáját sok helyen vadásszák, hazájában, az Ibériai-félszigeten fontos ökológiai szerepet betöltő, jelenleg veszélyeztetett, a 20. század első fele óta csökkenő populációjú faj (Marín-García & Llobat, 2021, IUCN Red List: <https://www.iucnredlist.org/species/41291/170619657>, 2025.05.09). Ezzel ellentétben számos helyen, ahol betelepített fajként van jelen, elszaporodott, kártevőként tartják számon, és a populáció szabályozása problémát jelent, így például Ausztráliában (Alves, J. M. *et al.*, 2022), Amerika bizonyos részein és számos szigeten (Bello-Rodríguez *et al.*, 2021; Delibes és Delibes-Mateos, 2015). A házinyulat hús, szőrme, illetve gyapjú termelési célból is tenyésztik, emellett változatos laboratóriumi célokra is használják. A széles körben kapható nyúl antitestek termeltetése mellett betegségmodellként is alkalmazzák, különösen szív- és érrendszeri problémák modellezésénél (Bósze *et al.*, 2016; Fan, J. *et al.*, 2018; Niimi *et al.*, 2020), de léteznek elsődlegesen házikedvencként tenyésztett fajták is.

Az *Oryctolagus* nemzetséggel szemben a *Lepus* nemzetség számos fajt foglal magában. A *Lepus* fajokat régóta vadásszák húsupért (Lloveras *et al.*, 2016), és bár az emberrel való kapcsolatra utaló nyomok több helyen megtalálhatóak (Sheng *et al.*, 2020; Somerville *et al.*, 2017), ma háziásított formájuk nem létezik. Ugyanakkor természetes módon előfordulnak a világ számos pontján, sok esetben több faj található egymással átfedő élőhelyeken, és számos esetben ismert közöttük hibridizáció előfordulása a régmúltban (Alves, P.C. *et al.*, 2003; Ferreira *et al.*, 2021) és ma is (Wu *et al.*, 2011; Jones *et al.*, 2018; Levänen *et al.*, 2018b). A *Lepus* fajok tehát a természetes úton történő hibridizációra, valamint annak hosszú- és rövid távú hatásainak megfigyelésére is kitűnő modellek, az élőhelyek változása, és az egyes életterekhez alkalmazkodott rokon fajok egymásra hatásának tükrében.

Az általam is vizsgált két *Lepus* faj közül a mezei nyúl (*Lepus europaeus*) (1. ábra) Eurázsia nyugati részében nagy területen elterjedt. Élőhelye elsősorban nyílt, füves területek, illetve mezőgazdasági területek (Bock, 2020). A havasi nyúl (*Lepus timidus*) (1. ábra) a mezei nyúlnál valamivel kisebb testű, hideg éghajlathoz alkalmazkodott faj. Észak-Európában és Szibériában elterjedt, élőhelye a tundra és tajga. Az IUCN információja (<https://www.iucnredlist.org/species/11791/451771981>) alapján jelenleg elismert 16 alfaj közül Finnország területén a nominális (*Lepus timidus timidus*) alfaj elterjedt. Az embert, szemben a mezei nyúllal, lehetőség szerint kerüli (Angerbjorn és Flux, 1995): míg a mezei nyúl ezeken a területeken az ember által befolyásolt, főként mezőgazdasági területeken él, de akár városokban is előfordul, a havasi nyúl jellemzően az erdős, ember által jobbra érintetlen területeket preferálja.

Ma a klímaváltozás, és a mezei nyúl ezáltal elősegített gyors terjedése észak felé Finnország területén egyaránt nyomást gyakorol a havasi nyúl populációra. E dolgozat írásakor a mezei nyulat már az északi sarkkörtől északra eső helyszíneken is megfigyelték. Az IUCN oldalán

(<https://www.iucnredlist.org/species/41280/45187424>) elérhető elterjedési térkép alapján a faj élőhelyének északi határa ennél még jelentősen délebbre húzódik. Ezzel párhuzamosan Finnország déli részén a havasi nyúl populációk csökkenő tendenciát mutatnak. Az élőhelyek és populációk felgyorsult változása konzervációs szempontból kiemelten fontossá teszi a megfelelő minőségű genomikai adatok elérhetőségét.



1. ábra – A) Mezei nyúl (*Lepus europaeus*) és B) havasi nyúl (*Lepus timidus*) egyedek. A havasi nyúl ikonikus jellemzője a teljesen fehér téli bunda. Fotók: Dr. Mervi Kunnasranta

3.3. Hibridizáció és szelekció

3.3.1. A domesztikáció hatásai

A domesztikáció során kiválasztott tulajdonságok, és ebből adódóan a szelekció hatása alatt álló gének számos háziasított fajban átfednek egymással. Ilyen tulajdonságok például a testméret, a növekedési erély, a szőr tulajdonságai, neurológiai- és viselkedési minták, úgymint az agresszió csökkenése, vagy az emberekkel szembeni nagyobb tolerancia. A kiválasztott tulajdonságok közös pontja, hogy alkalmasabbá teszik az állatokat az emberek közelében, az emberek által kontrollált környezetben való életre, valamint növelik azok gazdasági értelemben vett hasznosságát.

A nyúl domesztikációja a relatív közelmúltban történt, bármely pontot vesszük az időben a háziasítás kezdeteként. Ez lehet az első emberrel való asszociáció, vagy a szándékos tenyésztés és kiválasztás kezdete, de a nyúl háziasítása semmiképp sem kezdődött régebben, mint 1,5 – 2 ezer év (Alves, J. M. *et al.*, 2015). A házinyúl, ellentétben sok más háziasított állatfajjal, egy földrajzi területhez és egy eredeti fajhoz kapcsolható (Carneiro *et al.*, 2011). A házinyúl populáció mérete ennek következtében többször erősen összeesett, először a kezdeti domesztikációs lépéskor, majd a fajták kitenyésztésének kezdetekor, a vad üregi nyúlhoz képest jelentősen alacsonyabb genetikai diverzitást eredményezve (Alves, J. M. *et al.*, 2015).

Carneiro és munkatársai (Carneiro, *et al.*, 2014b) rámutattak a nyúl domesztikációjának sokgénese eredetére, amely egyezik számos más háziasított fajban találtakkal (Ramos-Onsins *et al.*, 2014; Rochus *et al.*, 2018; Heikkinen *et al.*, 2020; Zheng *et al.*, 2020; Lawal és Hanotte, 2021; Xia *et al.*, 2021). Ez alapján háziasítás folyamán bekövetkezett változások nagy valószínűséggel gyakrabban történnek az allélgyakoriság változása, mint új mutációk miatt (Carneiro *et al.*, 2014b). Kutatások során nagy mennyiségű gént és molekuláris útvonalat azonosítottak már, mint szelekció alatt állót, amelyek szerepet játszanak a nyúl szőrzete színének és struktúrájának kialakításában,

illetve testméretének befolyásolásában (Ballan *et al.*, 2022, 2023; Carneiro *et al.*, 2017; Demars *et al.*, 2018; Fatima *et al.*, 2023; Li, C. *et al.*, 2022; Liu, C. *et al.*, 2021). A háziasítás során számos esetben a változások az idegrendszert érintő, érzékszervi, vagy viselkedéssel kapcsolatos eltéréseket eredményeznek (Carneiro, *et al.*, 2014a). Ugyanakkor a szelekció hatására kialakuló több változás nem magukban a génekben történik hiszen a génexpresszió, és ebből adódóan a fenotípus megváltozása lehetséges a szabályozó régiókban bekövetkező változások hatására is (Sato *et al.*, 2020). A genetikai- és génexpressziós változások mellett a házinyúl agyának fiziológiai változásai is megfigyelhetők (Brusini *et al.*, 2018).

A háziállatok reprodukciós ciklusa is gyakran különbözik a vad változatokétól. Az ember által létrehozott, kontrollált környezet megbízhatóbb, konzisztens feltételeket teremt a szaporodáshoz, amely önmagában is befolyásolhatja azt. Üregi nyúlban először Carneiro *et al.* (2015) mutatta ki, hogy a házinyúl egész éves szaporodása legalább részben genetikai szelekció eredménye, de hasonló eredményeket találhatunk más háziállatokkal kapcsolatban is (Karlsson *et al.*, 2016; Signer-Hasler *et al.*, 2022).

3.3.2. A házi és vad populációk közötti hibridizáció lehetősége és potenciális hatásai

Annak ellenére, hogy számos tanulmány foglalkozik a házinyúl fajták közötti- és a fajtákon belüli szelekcióval, ezek nem vizsgálják azok kapcsolatát a vad populációkkal, illetve a domesztikáció átfogó, minden házinyúlra kiterjedő hatását sem. Emellett, bár az új változatoknak az egyes populációkra gyakorolt hatása a szabadon, illetve szabadabban hibridizálódó házi- és vad formában is létező fajokban is megfigyelhető (Besnier *et al.*, 2022; Heikkinen *et al.*, 2020; Pilot *et al.*, 2021), a vad- és házi üregi nyúl hibridizációjának ökológiai hatásai nem vizsgáltak alaposan.

Az üregi nyúl (*Oryctolagus cuniculus*) házi- és vad egyedei egymással hibridizációra képesek, köztük reprodukciós barriert eddig nem figyeltek meg. Ezzel szemben az Ibériai-félszigeten előforduló *O. c. cuniculus* és *O. c. algirus* alfaj között létezik ilyen korlát (Carneiro *et al.*, 2014a). A házi- és vad egyedek hibridizációja ökológiai problémákhoz vezethet a vad állomány genetikájának befolyásolása által. Más fajokon végzett tanulmányok két fő problémára világítanak rá a vad- és háziállatok hibridizációjával kapcsolatban. Egyrészt, a hibridek kiszoríthatják a vad populációkat (Smyser *et al.*, 2020), másrészt az előforduló hibridizáció mennyisége visszafordíthatatlanul megváltoztathatja a vad állomány genetikai összetételét (Grossen *et al.*, 2014; Quilodrán *et al.*, 2020; Wu *et al.*, 2020). Léteznek példák olyan esetekre is, ahol a háziasított állatok tulajdonságaira pozitív szelekció történik a vad populációkban, amely hozzájárul azok túléléséhez (Grossen *et al.*, 2014; Pilot *et al.*, 2021).

3.3.3. Hibridizáció és introgresszió a természetben

A fajok közötti introgresszió számos esetben visz be olyan új tulajdonságokat, amelyek pozitív szelekciója új változatok által szelekciós előnyt biztosít a másik fajnak.

Több tanulmány foglalkozik Leporidae fajok közötti hibridizációval, introgresszióval és ezek következményeivel (Alves, P. C. *et al.*, 2003; Ferreira *et al.*, 2021; Jones *et al.*, 2018; Levänen *et al.*, 2018a, 2018b; Melo-Ferreira *et al.*, 2009; Wu *et al.*, 2011). A fajok közötti reprodukciós korlát jelenléte csökkenti a hibridizáció során bekerülő gének megmaradásának esélyét: amennyiben a génekre a hátrányos változatokkal szembeni negatív szelekció (purifying selection) hat, annak következtében néhány generáción belül eltűnhetnek a hibridizáció nyomai. Ahhoz, hogy az ilyen módon kapott gének a populációban elterjedjenek, és tartósan megmaradjanak, elvileg elegendő

lehet a genetikai drift és a véletlen is, de vannak esetek, ahol bizonyított, hogy az introgresszió valamilyen evolúciós előnyt jelent a fogadó fajban. *Lepus* fajok esetében ilyenek lehetnek a téli környezethez megfelelő szőrszín kialakításában (Giska *et al.*, 2022; Jones *et al.*, 2018) szerepet játszó gének.

A fajok közötti természetes hibridizáció és genetikai introgresszió sokkal nagyobb hatású és általánosabb jelenség, mint azt a múltban feltételezték. A detektálási módszerek fejlődésével nem csak számos fajról derült ki, hogy természetes módon hibridizálódni képes másokkal, de annak ökológiai és evolúciós szerepe is egyre jobban vizsgált, és nem elhanyagolható (Zbinden *et al.*, 2023). A hibridizációt gyakran tekintik az érintett fajok és a biodiverzitás számára alapvetően hátrányosnak (Rhymer és Simberloff, 1996; Todesco *et al.*, 2016), de egyre több tanulmány igazolja annak lehetséges pozitív hatásait, adaptív allélok introgressziója által. Ritka esetben a hibridekből akár teljesen új fajok is létrejöhetnek (Taylor és Larson, 2019).

Az egyes fajok elterjedési területének szélén általában aránylag kevés egyed képviseli az adott fajt, és a populáció sűrűsége is alacsonyabb a fő elterjedési területen tapasztaltnál. Ez hatással van az itt található egyedek genetikai diverzitására: az alapító hatás (founder effect) és a genetikai drift együttes hatása könnyen a sokszínűség csökkenéséhez vezethet (Peischl *et al.*, 2015). Ezzel szemben az ugyanezen területen megtalálható rokon fajjal történő hibridizáció új genetikai változatokat vihet be, és az azokkal létrejövő új kombinációk nem csak a populáció genetikai diverzitását növelik, de a környezethez való alkalmazkodást is segíthetik (Glasenapp és Pogson, 2024; Jones *et al.*, 2018; Yao *et al.*, 2019).

Általánosságban véve minél fiatalabb evolúciósan két közeli rokon faj, annál kisebb mértékű, illetve könnyebben átjárható a reprodukciós korlát. A dolgozatomban vizsgált *Lepus timidus*, és annak közeli rokon faja, a *Lepus europaeus* kb. 5 millió éve (Pohjoismäki *et al.*, 2021), evolúciós léptékben nem sok ideje vált el.

A természetes úton végbemenő, fajok között hibridizáció hatásai nagyon hasonlóak is lehetnek a házi- és vad változatok közötti hibridizációjához. Ebben az esetben is előfordulhat, hogy a hibridek jelenléte véglegesen megváltoztatja a genetikai állományt, vagy azok kiszorítják az eredeti fajokat. Gyengébb alkalmazkodó képességű, de nagy mennyiségű hibrid esetén pedig a populáció növekedését gátolhatják (Todesco *et al.*, 2016). Az introgresszió nem feltétlenül azonos mértékben hat a hibridizálódó fajok mindegyikére, az introgresszió gyakran tolódik el a hibridizálódó fajok közül az egyik irányába (Currat *et al.*, 2008). *Lepus* fajok esetében például gyakoribb a havasi nyúl (*Lepus timidus*) felől a mezei nyúl (*Lepus europaeus*) irányába történő introgresszió, mint ellenkezőleg (Levänen *et al.*, 2018b; Pohjoismäki *et al.*, 2021).

Az autoszómákon megfigyelhető introgresszió mellett a hibridizálódó populációkban gyakori a mitokondriális introgresszió is. Mitokondriális introgresszió alatt általában a teljes mitokondriális DNS introgresszióját értjük. A finnországi *Lepus* hibrid zóna területén az mtDNS introgressziójának számos esetét megfigyelték (Levänen *et al.*, 2018a; Tapanainen *et al.*, 2024).

3.3.4. A hibridizáció genomi lenyomatai

Bár a hibridizáció elvileg eltörölhetné a populációk közötti határokat az azokban jelenlévő genetikai variációkat mindegyikben elterjesztve, ez nem feltétlenül így történik. A Finnországban ma egyszerre jelenlévő, és generációk óta hibridizálódó *Lepus europaeus* és *Lepus timidus* például nem mutatja populációik összeolvadásának jeleit (Levänen *et al.*, 2018a). Megfelelő körülmények

között az előnyös változatok introgressziója akár fel is gyorsíthatja a speciációs, illetve kolonizációs folyamatokat (Hamilton és Miller, 2016). A populációk, fajok és változatok közötti határok ugyanakkor genomikai szempontból általában valamelyest elmosódottá válnak (Milián-García *et al.*, 2015).

Az introgresszált szakaszok, és ebből adódóan az erre utaló genomi jelek egyenetlenül oszlanak el a genomon. Az introgresszióra a drift, pozitív és negatív szelekció egyaránt hatással vannak (Hedrick, 2013; Payseur és Rieseberg, 2016).

Az egyes elkülönülő populációk, genetikai származási vonalak közötti határok különböző mértékben lehetnek átjárhatók. Közeli rokon fajok esetében például azok termékeny hibrid utódok létrehozására is képesek lehetnek, de változó mértékű introgressziót mutatnak a genom különböző régióiban. Hasonló jelenség figyelhető meg az üregi nyúl két alfaja között is (Carneiro *et al.*, 2014a).

A hibridizálódó populációk között megosztott, azaz mindegyikben előforduló allélek vizsgálatok az introgresszióhoz hasonló genomi szignált okozhat egy-egy ősi közös (shared ancestral) allél, amely a fajképzés, vagy domesztikáció során is megmaradt (Wiuf *et al.*, 2004). Így a hibridizáció vizsgálatához fontos, hogy rendelkezésünkre álljanak nagy biztonsággal tisztán az eredeti populációk egyedeinek tekinthető állatok.

A mitokondriális introgresszió azonosítása egyszerűbb az autoszomálisnál, és előfordulása a hibridizáció egyértelmű jele. A mitokondriális DNS (mtDNS) emlősökben uniparentális, azaz egy szülő, emlősök esetében még pontosabban maternális (anyai) öröklődése az esetek legnagyobb részében homoplazmikus utódokhoz vezet, azaz egy utódban egy haplotípus lesz megtalálható. Ennek köszönhetően a rekombináció lehetősége általában elhanyagolható, és a haplotípusok akár 1-2 tipikus genetikai változatot felhasználva azonosíthatóak. Nem elhanyagolható azonban, hogy a finnországi *Lepus* hibridizációs zóna területén például aránylag nagy számban fordulnak elő LR régióra heteroplazmikus egyedek (Tapanainen *et al.*, 2024). Az introgresszált mtDNS pedig generációkon át megmaradhat, akár olyan területeken is, ahonnan a faj azóta eltűnt (Alves, P. C. *et al.*, 2008; Thulin *et al.*, 1997; Thulin és Tegelström, 2002)

3.3.5. A hibridizáció detektálása

A hibridizáció mértékének vizsgálatához általában szükségünk van előzetes feltételezésre a felmenő populációkba tartozó egyedekről. A nem-hibrid szülő populációk meghatározása általában területi, morfológiai jellemzőkön alapul. Ezzel kerülnek összehasonlításra az ismeretlen hovatartozású, vizsgált egyedek, amelyekben célunk megbecsülni az egyes populációkból örökölt genomi szekvenciák mértékét. Minél jobb háttértudással rendelkezünk a nem-hibrid populációkról, minél több a vizsgált marker, annál jobban becsülhető az egyes egyedek származása.

A genomi introgresszó vizsgálata esetében az egyes genomi szakaszok eredetének lokális vizsgálatát végezzük. Ekkor a teljes genom egyes szegmenseinek vizsgáljuk az eredetét (Browning *et al.*, 2023; Guan, 2014; Sankararaman *et al.*, 2008; Tang *et al.*, 2006). *Lepus* fajokban például nagy felbontású, WGS alapú vizsgálatot végzett Souto *et al.* (2025), több ibériai-félszigeti faj populációjának evolúciós történetének és genetikai örökségének felderítésére. Amennyiben nem szükséges nagy felbontású vizsgálat, akár néhány marker segítségével is kimutatható a hibridizáció

következtében bekövetkezett introgresszió: Pohjoismäki *et al.* (2021) az *L. timidus* és *L. europaeus* között az MHC régióban 3 marker segítségével vizsgálta azt.

Globális vizsgálatot végezhetünk, ha az introgresszió lokalizációja nélkül a hibridizáció általános mértékét célunk felderíteni. A közelmúltban történt, vagy nagyobb mértékű hibridizáció esetében idegen populációból származó nagyobb mennyiségű allél előfordulása várható a hibridekben. A globális becslésnek elterjedt módszere a STRUCTURE (Pritchard *et al.*, 2000) és ADMIXTURE (Alexander *et al.*, 2009) programok. A hasonló elven alapuló, modell alapú algoritmusok közös vonása, hogy a minták probabilisztikusan kerülnek populációkba sorolásra, egy, vagy egyszerre akár több populációba tartozóként is. A kimenet mindkét esetben százalékos becslést ad az egyes populációkba tartozásról. A modell alapú módszerek alapfeltételezése, hogy az egyes csoportokba tartozó megfigyelések egy parametrikus modellből származó random minták. Az egyes csoportok tulajdonságainak, és az egyedek egyes csoportokhoz való tartozásának meghatározása egymáshoz kapcsoltn történik. A két algoritmus precizitása és pontossága hasonló, de közülük az ADMIXTURE hatékonyabb nagy mennyiségű adat kezelésében (Alexander *et al.*, 2009).

Az EIGENSTRAT (Patterson *et al.*, 2006; Price *et al.*, 2006) ezzel szemben főkomponens-analízisre alapul, eredményei pedig nem százalékban kifejezett asszociációk az egyes felmenő csoportokkal.

Az ADMIXTURE és a STRUCTURE hosszú ideje töretlen népszerűsége mellett is, a populációk és egyedek hibrid státuszának meghatározására szolgáló módszerek folyamatosan fejlődnek. A gyakran használt Patterson-féle D-statisztikára (Durand *et al.*, 2011; Patterson *et al.*, 2012) alapuló, annak erejét felülmúló új statisztikák jelentek meg az elmúlt néhány évben, mint a Cauchy combination teszt (Liu és Xie, 2020), a MinP (Chen, 2022), vagy Kubatko és Chifman módszere (Kong és Kubatko, 2021; Kubatko és Chifman, 2019).

3.3.6. A szelekció genomi lenyomatai

Amennyiben egy új allél adaptációs előnyt jelent, az elvileg az adott allél túlreprezentáltságához (overrepresentation) vezet a populációban annak pozitív szelekciója által. Ez könnyen megfigyelhető az egyértelmű fenotípussal rendelkező gének esetében, mint amilyenek a téli szőrszint befolyásoló gének is, a *Lepus americanus* (Jones *et al.*, 2018, 2020a, 2020b), vagy a *Lepus timidus* (Ferreira *et al.*, 2021; Giska *et al.*, 2019) esetében is. A téli fehér szőrszín megléte vagy hiánya nem csak könnyen megfigyelhető, de evolúciós előnye is meglehetősen nyilvánvaló a hóborította területeken.

A szelekció minden esetben nyomot hagy a genomon, akár természetes, akár ember által irányított változásokról van szó – az minden esetben a környezet változásaira és a genetikai mutációkra adott válasz, amelynek lenyomata detektálható. A genetikai szelekcióra utaló általános jelek közé tartozik adott genomi régióban, illetve a szelektált lókusznál környékén az alacsonyabb nukleotid diverzitás, a populáción belüli hosszú, illetve nagy gyakoriságú haplotípusok, több populáció összehasonlítása esetén pedig az azok közötti magasabb differenciációs értékek is.

Egy adott változatra történő szelekció lehet pozitív – előnyös variáció – vagy negatív – hátrányos variáció – is. A kérdéses genetikai variáció lehet új mutáció vagy létező változat is, amely valamilyen környezeti tényező változására válik előnyössé vagy hátrányossá. A pozitív szelekció várhatóan egy allél frekvenciájának növekedéséhez vezet a populációban, a negatív annak csökkenéséhez. Speciális eset a kiegyensúlyozó szelekció (balancing selection), amikor a

heterozigóta egyedek rendelkeznek szelekciós előnnyel, így az allélfrekvencia körülbelül állandó marad – ennek tipikus példája a sarlósejtes vérszegénység (Rees *et al.*, 2010) –, illetve a diverzifikáló (diversifying/disrupting) szelekció, amely során eltérő extrém fenotípusok egyaránt preferáltak (Nielsen, 2005).

Típusától függően a szelekció változó jeleket hagyhat a genomon. Az úgynevezett „selective sweep”-ek, amelyek során az előnyös variáció frekvenciája megnövekszik a populációban, intenzitástól függően lehetnek „hard” vagy „soft sweep”-ek. Előbbi esetben egy új mutáció frekvenciája gyors növekedést mutatva hamar eljut az új allél teljes, vagy közel teljes fixációjáig, míg utóbbi esetben több, az előnyös allélt tartalmazó haplotípus egyszerre kerül kiválasztásra. Ebben az esetben a variáció már létezett korábban, a rekombináció és más mutációk is befolyásolták annak környezetét a genomon, mielőtt az valamilyen külső környezeti változás hatására előnyössé vált. A „soft sweep” egy tág fogalom: a szelekció több típusa is hasonló nyomot hagy a genomon (Hermisson és Pennings, 2017; Pennings és Hermisson, 2006).

A sokgénes tulajdonságokra történő szelekció detektálását bonyolítja, hogy számos, kis hatású allél vesz benne részt. Ezek gyakorisága legnagyobb valószínűséggel enyhe növekedést fog mutatni a populációban, de egy allél sem fog teljes fixációt elérni (Berg és Coop, 2014; Fan, S. *et al.*, 2016).

A szelekció detektálását nehezíti a genetikai drift, amelynek köszönhetően egyes semleges változatok a véletlennek köszönhetően terjedhetnek el, vagy akár tűnhetnek el a genomból. A demográfiai folyamatok és a rekombináció hatásai sem elhanyagolhatóak. Gyorsan növekvő populációk esetében a genetikai változatosság elmaradhat a populáció mérete alapján várttól, míg a gyors csökkenésnek ellenkező hatása van. Így demográfiai okokból is kialakulhatnak a szelekcióéra hasonlító lenyomatok a genomon (Crisci *et al.*, 2013; Provine, 2004), vagy a valójában végbement szelekciós folyamatától eltérőre utaló jeleket kaphatunk (Wilson *et al.*, 2014).

3.3.7. A szelekció detektálása

A szelekciós szignálok detektálására a relatív egyszerű statisztikáktól komplex programokig számos módszer létezik, amelyek a genetikai változatok detektálásának hatékonyságával és a számítási kapacitás növekedésével együtt fejlődtek és fejlődnek ma is. Az első statisztikai módszerek fejlesztett és korrigált változatai képezik a mai, összetett detektáló programok alapját is. A szelektált régiók detektálására létező módszerek néhány alapvető kategóriára oszthatók fel.

Kapcsoltsági egyensúlytalanságon (linkage disequilibrium, LD) alapuló módszerek azon feltételezés alapján detektálják a kiválasztásra került régiókat, hogy a szelektált lokusz körül hitchhiking történik: az előnyös változattal párhuzamosan a környező, semleges variációk frekvenciája is megnő (Barton, 1998; J. M. Smith és Haigh, 1974). Ilyen módszer például Jacobs *et al.* (2016) statisztikai módszere. Bizonyos esetekben, „hard sweep”-ek esetében a „hitchhiking” rekombinációs hotspotokhoz hasonló mintázatot eredményezhet (Kim és Nielsen, 2004). A rekombinációs térképek használata ellensúlyozhatja ennek hatását, ugyanakkor a befolyás kétirányú: a szelekció és a rekombináció jelei kölcsönösen elfedhetik egymást (Horscroft *et al.*, 2019).

A haplotípus alapú módszerek alapelve nagyon hasonló az LD alapú módszerekéhez. Eszerint a populációban magas frekvenciával előforduló, hosszabb haplotípusok a hitchhikingnek köszönhetőek. A haplotípusok frekvenciájára (Garud *et al.*, 2015) és hosszára (Ferrer-Admetlla *et*

al., 2014), valamint ezek kombinációjára is alapulnak módszerek (Carvajal-Rodríguez, 2017; Lange és Pool, 2016).

A legutóbbi közös ős óta eltelt időre (time to most recent common ancestor, TMRCA) alapuló módszerek a várttól eltérő mintázatokat keresnek a genomon, a heterozigóta és homozigóta változatok számát összehasonlítva. A heterozigóta lókuszok magasabb száma hosszabb TMRCA-ra utal. A kalkuláció egy módszerét (Li és Durbin, 2011) írja le, míg a TSEL (Hunter-Zinck és Clark, 2015) a módszer egy applikációja.

A composite likelihood alapú módszerek alapvető feltételezése, hogy az allél frekvencia spektrum torzulni fog a szelektált környéken a genom többi részéhez képest. Egy szelektált lókusztól más, olyan lókuszokkal lesz körülvéve, amelyeken az azzal asszociált allélek frekvenciája szintén magasabb – vagy alacsonyabb, a szelekció irányától függően. A „composite likelihood ratio” módszerek alapja Kim és Stephan (2002) munkája, amelyet módosítva több szoftver is felhasznál, például a SweeD (Pavlidis *et al.*, 2013) és a SweepFinder2 (DeGiorgio *et al.*, 2016) is. Több teszt párhuzamos használatával azok eredményei megerősíthetők, vagy megcáfolhatják egymást. Több tesztelési módszert integráló megoldásokat publikált például Ma *et al.* (2015) és Randhawa *et al.* (2014).

A gépi tanulási módszerek elterjedésével azok a populációgenetikában, illetve a szelekció detektálásában is megjelentek, ilyenek például Pybus *et al.* (2015), Schrider és Kern (2016), valamint Sheehan és Song (2016) munkái. Populációgenetikában alkalmazott gépi tanulási módszerekről részletesebb áttekintést ad Huang *et al.* (2024).

A detektálás legnagyobb problémái közé tartozik – a biológiai háttértényezőknél, azaz a demográfiai változásokon és a rekombináción túl –, hogy a genetikai variációknak nincs ismert semleges eloszlása, amely alapján a statisztikai szignifikancia meghatározható lenne (Kemper *et al.*, 2014). Emiatt számos esetben egyszerűen olyan szakaszokat tekintünk szelektátnak, amelyek az eredmények extrém értékei közé tartoznak (pl.: Ballan *et al.*, 2022; Papachristou *et al.*, 2020; Rostamzadeh Mahdabi *et al.*, 2021; Xia *et al.*, 2021).

A szelektív sweep-ek felismerésének módszereiről több áttekintő cikk jelent meg (Horscroft *et al.*, 2019; Huang *et al.*, 2024; Panigrahi *et al.*, 2023; Pavlidis és Alachiotis, 2017), és a szelekció háziállatokban történő detektálásával is számos publikáció foglalkozik szarvasmarha (Papachristou *et al.*, 2020; Randhawa *et al.*, 2014), házityúk (Rostamzadeh Mahdabi *et al.*, 2021), kecske (Signer-Hasler *et al.*, 2022; Zheng *et al.*, 2020), juh (Rochus *et al.*, 2018), nyúl (Ballan *et al.*, 2022, 2023; Carneiro *et al.*, 2015), és más fajok esetében is.

3.4. A rekombináció evolúciós szerepe és azonosítási módszerei

Az evolúció és szelekció egyik legfontosabb mechanizmusa a rekombináció. A rekombináció teszi lehetővé a fentebb tárgyalt genetikai introgressziót, meglévő genetikai változatok új kombinációban való megjelenését, ezzel új tulajdonságokat, jobb alkalmazkodási képességet létrehozva. A populációk hosszú távú túlélése múlik a környezeti változásokhoz való alkalmazkodási képességükön, az ehhez szükséges genetikai variáció megteremtői pedig a mutációk és a rekombináció.

Míg a mutációk az új allélek létrejöttéért felelősek, azok elterjedése a rekombinációnak köszönhető, azáltal, hogy lehetővé teszi az előnyös új változatok elválasztását a hátrányosaktól.

Mivel az új mutációk csupán kis hányada előnyös, azok a rekombináció mechanizmusa nélkül aligha maradhatnának meg és terjedhetnének el (Radman *et al.*, 1999).

A rekombináció és a speciáció is összefüggő folyamatok. A speciáció során a genom egyes régióiban csökkent rekombinációs ráta figyelhető meg. A közeli rokon fajok közötti kromoszóma-átrendeződések létrejötte nem feltétlenül a fajok kialakulásának része, de létezésük hozzájárul a rekombináció lehetőségének csökkenéséhez, és ezzel a kialakult fajok közötti határok fenntartásához (Butlin, 2005). A kromoszóma-átrendeződések befolyása a rekombinációra többrétű. Egyrésztől átrendezhetik a kapcsoltsági viszonyokat, másrésztől pedig megelőzhetik a rekombinációt heterozigóta egyedekben (Navarro és Barton, 2003). Nagy méretű, kromoszóma-szintű genomi átrendeződések fajon belül ritkán polimorfak (Butlin, 2005).

A rekombinációs ráták és mintázatok a legtöbb fajban nemtől is függenek. A hím és nőstény egyedek általában eltérő rátát és lokalizációt mutatnak. Az átkereszteződés hím egyedekben jellemzően a telomer régiókban, míg nőstényekben egyenletes eloszlású, vagy a centromer környékén gyakoribb (Sardell és Kirkpatrick, 2020). A változó rekombinációs ráták okaival, meghatározásával és elemzésével részletesen foglalkozik Peñalba és Wolf (2020).

3.4.1. A rekombináció mechanizmusai

A DNS kettős száltörései javítás hiányában végzetes következménnyel járhatnak a sejtre nézve. Az ilyen törések két fő lehetséges javítási útvonala a nem-homológ végek összekapcsolása (non-homologous end-joining, NHEJ) és a homológ rekombináció (HR). Az alkalmazott javítási mechanizmus a sejtciklus fázisától, és az eltört DNS végének típusától függ (Symington és Gautier, 2011).

A homológ rekombináció a sejtciklus S és G2 fázisában preferált, amikor elérhetőek az ahhoz szükséges homológ szekvencia templátok is. Míg a NHEJ során egyszerűen a törött végek ligációja történik, amely során nagy a hibalehetőség, a HR-hez szükséges egy szekvencia templát, amelynek segítségével kiküszöbölhető az információ-vesztés. A HR egyben lehetővé teszi a replikációs villa fennakadásából adódó problémák kezelését, javítását is (Elbakry és Löbrich, 2021).

A HR annak kezdeti lépéseit követően több mechanizmus útján haladhat tovább. Az első lépés a törések végének módosítása: a nukleázok 3' egyszálú DNS „overhang”-et, túllógó szekvenciát generálnak, amit amelyet az RPA (replication protein A) elfed. Az RPA-t rekombináz enzim váltja le, amely nukleoprotein szálakat képez, megindul a homológia keresés. A komplementer szekvencia megtalálását követően, létrejön a D-loop (displacement loop). A primer és a templát összekapcsolódásával megindul a szintézis útján történő javítás. Ennek megkezdését követően három szubmechanizmussal folytatódhat a folyamat.

Ezek egyike a Holliday-junction-ökön (HJ) keresztül történik. A HJ egy négyirányú kapcsolat a rekombinációban részt vevő szálak között (Holliday, 1964). Az összefonódó szálak elválasztása két mechanizmuson keresztül történhet. Az egyik során a szálak elválasztásakor a rekombinálódó molekulák genetikai információcsere nélkül válnak el egymástól (Wu és Hickson, 2003). A második esetben struktúrára szelektív nukleázok segítségével nagyjából egyenlő arányban keletkeznek crossover és non-crossover termékek (Elbakry és Löbrich, 2021). A kettős HJ gyakori intermediere a homológ rekombinációnak.

A „synthesis-dependent strand annealing” (SDSA) során Holliday-junction intermedierek létrejötte nélkül disszociálnak az új szálak a templáttól, crossover termékek létrejötte nélkül (Miura *et al.*, 2012).

A kettős-száltörés indukált replikáció (break-induced replication, BIR) esetében hosszú távú szintézis indul meg az inváziós DSB szálvégről, a másik vég bekapcsolódása nélkül. A BIR gyakran vezet mutációkhoz, akár nagy léptékű szekvencia-átrendeződésekhez is (Kramara *et al.*, 2018).

3.4.2. Rekombináció azonosítása a genomon

A rekombinációs intermedierek létrejötte megfigyelhető transzmissziós elektron-mikroszkópia (TEM) (Thresher és Griffith, 1992) vagy Southern blot (Codner *et al.*, 2021; Li, D. *et al.*, 2019) segítségével. Rekombináció követésére létrehoztak modell szervezeteket is (Hendricks és Engelward, 2004). A rekombinációs szekvenciák azonosítása, a rekombináció pontos következményei pedig szekvenálási, bioinformatikai módszerekkel deríthetők fel.

Rövid leolvasásos szekvenálások esetében a rekombinációs események felderítését valószínűségi számítás alapú módszerek támogatják. Ezek a módszerek olyan esetekben működnek jól, ahol a rekombináció gyakran megfigyelt, jól ismert jelenség.

A statisztikai módszerek két fő csoportra oszthatók, a szubsztitúció-disztribúciós módszerekre, amelyek a szubsztitúciók csoportosulását tesztelik gén szekvenciákon, és a filogenetikai módszerekre, amelyek a genomon szomszédosan elhelyezkedő szekvenciák alapján készült filogenetikai fák közötti eltérések alapján azonosítják a rekombinációt (Bay és Bielawski, 2011). Filogenetikai alapú módszer például a BARCE (Husmeier, 2003), a DualBrothers (Suchard *et al.*, 2002), az RDP (Martin és Rybicki, 2000), és további generációi, egészen az RDP4-ig (Martin *et al.*, 2017), a GARD (Kosakovsky Pond *et al.*, 2006) és az arra építő további megoldások is. Szubsztitúció disztribúción alapul például a GENECONV, MaxChi, és a Chimaera (Bay és Bielawski, 2011), és a ReCombine programcsomag (Anderson *et al.*, 2011).

A genotipizálási hibák befolyásolhatják a tesztek eredményét (Saunders *et al.*, 2007). Ennek kiküszöbölésére törekszik például a haploMAGIC (Montero-Tena *et al.*, 2024) amely elsősorban sokszülős (multiparentális) növényi populációkban való alkalmazásra lett kifejlesztve.

A rekombinációs események a genom bizonyos részein, a rekombinációs hotspotokon nagyobb sűrűséggel fordulnak elő. A rekombináció következményeinek további vizsgálata nélkül a hotspotok detektálására is léteznek eszközök (Paul *et al.*, 2016).

Az említett, számos rendelkezésre álló módszer azonban a legtöbb esetben a meiotikus rekombinációra fókuszál, vagy olyan prokarióta, illetve virális szekvenciákra, amelyek esetén a rekombináció létezése általánosan elfogadott.

3.4.3. Mitokondriális rekombináció

A mitokondriális rekombináció létezése emlősökben sokáig vitatott volt, annak ellenére, hogy az élővilág más területein a jelenség jól ismert, és sokszor megfigyelt. Puhatestűekben, ahol az mtDNS biparentális öröklődésű, bizonyítékát találták evolúciós időskálán történt mitokondriális rekombinációnak (Ladoukakis *et al.*, 2011; Ladoukakis és Zouros, 2001). Növényekben és élesztőkben pedig ma már nem csak elfogadott a rekombináció e specifikus esetének létezése, de az azt irányító enzimek is ismertek (Ling, 2002; Shedje *et al.*, 2007).

Mitokondriális rekombináció előfordulása emlősökben ugyan dokumentált ritka esetekben, például humán heteroplazmia esetében (Kraytsberg *et al.*, 2004; Zsurka *et al.*, 2005, 2007), az nem vált széles körben elfogadottá. Az emlős mtDNS rekombináció létezésének bizonyítását célzó korábbi kísérletek egyik sokat kritizált pontja volt, hogy azok rövid leolvasásos szekvenálásokat használtak fel, amelyek a nukleáris genomhoz képest rövid mitokondriális genomnak – gerincesek

esetében kb. 14-20 kb (Montaña-Lozano *et al.*, 2022), jelentősen rövidebb a növényi, vagy élesztő mitokondriális genomnál is – is csak töredékét fedik le egy leolvasással. A rövid leolvasások összeszerelése, egymáshoz való viszonya a rövid átfedő szakaszok miatt mindig bizonytalanabb, mint hosszú leolvasások esetében. A rövid leolvasásos szekvenálási technológiák ezen felül általában amplifikációs lépéseket is tartalmaznak, amely egyes szekvenciák relatív dúsulásához vezethet, csökkentve más szekvenciák detektálásának esélyét, valamint kimérikus szekvenciákat is generálhatnak (Hagstrom *et al.*, 2014), amelyek fals rekombinánsként jelenhetnek meg az analízis során.

Az eredmények bizonytalanságát mutatja, hogy a mitokondriális rekombináció fentebb említett megfigyeléseivel (Kraytsberg *et al.*, 2004; Zsurka *et al.*, 2005, 2007) ellentétben humán cybrid sejteken végzett kísérletek során Sato és munkatársai (2005) nem találták bizonyítékát rekombinációnak, és heteroplazmikus egér sejtvonalak esetében is csak három potenciálisan rekombináns mitokondriális genomot sikerült detektálniuk.

Az úgynevezett „copy-choice” rekombinációs mechanizmus mikroholog szekvenciák közötti templát cserére alapul (template switching). Mechanikailag hasonló a nukleuszban replikációs stressz hatására végbemenő folyamatokhoz, és könnyen vezethet az mtDNS integritásának problémáihoz (Persson *et al.*, 2019; Phillips *et al.*, 2017). Fragkoulis és munkatársai (2024) feltételezése alapján ez lehet emlősökben a mitokondriális rekombináció fő mechanizmusa, amelyet korábbi kísérletek is alátámasztanak (Persson *et al.*, 2019; Phillips *et al.*, 2017).

A mitokondrium és a mitokondriális genom egyedi jellemzői limitálják az eukarióta nukleáris genom rekombinációjának felderítésére alkalmazott genomikai módszerek használhatóságát a mitokondriális rekombináció vizsgálatában korlátozott. Az mtDNS prokarióta eredetének ellenére a bakteriális genomok vizsgálatára alkalmazott, elsősorban statisztikai módszerek (áttekintésük: Kalinina *et al.*, 2016) használhatósága is erősen limitált emlősök esetén. Ennek oka nem csak az, hogy a természetes módon előforduló rekombináció megfigyeléséhez heteroplazmikus egyedekre lenne szükség, hanem a normál sejtműködés során a mitokondriális rekombinációért felelős mechanizmusok elnyomása a lineáris mtDNS degradációja által (Peeva *et al.*, 2018).

3.5. A genomikai vizsgálatok háttérének megteremtése - A referenciagenomok

A genomikai kutatások fontos alapkövét képezik a referenciagenomok. A referenciagenom egy adott faj teljes genomjának szekvenciáját reprezentálja – legalábbis elméletben. A kisebb változatok, néhány nukleotidos rövid inszerciók és deléciók (indelek), illetve SNP-k mellett egy fajon belül is előfordulhatnak strukturális változások, nagyobb átrendeződések, akár olyan gének is, amelyek azonos faj más példányában nem, vagy több kópiában fordulnak elő (Gerdol *et al.*, 2020). A referenciagenomok célja kettős: a genomikai és genetikai analízisek egységes háttérének biztosítása, és ezzel együtt a genomai adatok feldolgozásának gyorsítása (Kaye és Wasserman, 2021).

Az elmúlt évek során az elérhető referenciagenomok mennyisége és minősége is megugrott, az elérhető szekvenálási technológiák és összeszerelési módszerek párhuzamos fejlődésének köszönhetően.

A fajkonzervációs erőfeszítésekben egyre fontosabb szerepet kapnak a megfelelő minőségű genomai adatok. A referenciagenomok fontosságát mutatja, hogy ma azok létrehozására számos, sokszor összefonódó konzorcium létezik, mint például a Darwin Tree of Life (DToL, <https://www.darwintreeoflife.org/>), az Earth Biogenome Project (Lewin *et al.*, 2018), a Vertebrate Genome Project (VGP <https://vertebrategenomesproject.org/>), European Reference Genome Atlas

(ERGA, <https://www.erga-biodiversity.eu/>), a Project Psyche (<https://www.projectpsyche.org/>), az Aquatic Symbiotic Genomics Project (ASG, <https://portal.aquaticsymbiosisgenomics.org/>), stb. A cél nem csak a referenciagenomok létrehozása, hanem egy azokhoz kapcsolódó minőségi sztenderd megalapozása és fenntartása (Rhie *et al.*, 2020a).

Rekombináció azonosítása során, vagy két, egymáshoz nagyon közel álló, hibridizálódó faj esetén az allélek megkülönböztetőségéhez különösen fontos a pontos referenciagenom, amely alapján a genomi variációk meghatározhatók. A *Lepus* nemzetség tagjai közül sokáig egy faj sem, majd néhány évig csak a *Lepus timidus* rendelkezett összeszerelt genommal, amely egy draft szekvencia volt, scaffoldingjához pedig az üregi nyúl referenciagenomot vették alapul (Marques *et al.*, 2020). A scaffold a contigokból összeállított, nem folytonos, azaz gap-ek (rések, ismeretlen szekvenciájú szakaszok) által elválasztott contigokat jelent, amelyekről kiegészítő információk alapján – például Hi-C kapcsolatok, vagy közeli rokon faj kromoszómaszerkezete – feltételezzük, hogy egymáshoz közel helyezkednek el, sok esetben ismert távolságra. Amellett, hogy az OryCun2 üregi nyúl referencia minősége elmarad a mai sztenderdektől, az *Oryctolagus* nemzetségbe tarozó üregi nyúl kariotípusa is eltér a *Lepus* nemzetségétől. Ezutóbbiak haploid genomja kettővel több autoszómát (23) tartalmaz a nyúlénál (21).

3.5.1. Biológiai kihívások a referenciagenomok összeszerelésében

A legtöbb nagy méretű genom erősen repetitív. Az ezáltal okozott nehézséget jól szemlélteti a repetitív szakaszok kitartó problémája a humán genomban (Chaisson *et al.*, 2015). A *Lepus europaeus* és *Lepus timidus* fajok például 46,8% (Michell *et al.*, 2024) és 42,35% (Fekete *et al.*, 2025a) százalékban tartalmaz repetitív szekvenciákat. A repetitív régiók problémájára a legkézenfekvőbb megoldást a kellően hosszú leolvasások alkalmazása jelenti. A „kellően hosszú” a régió hosszúságától függ: a leolvasás két vége nem-repetitív szakaszhoz kell, hogy illeszkedjen, olyan hosszúságban, amely egyértelművé teszi annak pozícióját. Az összeszerelt genomból hiányzó repetitív régiók hiányzó génekhez vezethetnek (Peona *et al.*, 2018).

A repetitív szakaszokon belül speciális problémát vetnek fel a centromer és telomer szakaszok. A centromer nehezen áthidalható lehet még modern hosszú leolvasásos technológiákkal is. A kromoszóma végén elhelyezkedő telomer szakaszok esetében azok természetéből adódóan nem lehetséges egy hosszú leolvasás mindkét oldali „horgonyzása” nem repetitív szekvenciákhoz.

Az ivari kromoszómák összeszerelésének kihívása szintén kapcsolódik a repetitív szakaszok jelenlétéhez. Egyes ivari kromoszómák mindössze egy rövid, pszeudo-autoszómális szakaszból és nagy mennyiségű repetitív szekvenciából állnak (Kejnovsky *et al.*, 2009; K. D. Smith *et al.*, 1987). Annak ellenére, hogy emlősökben a ivari kromoszómák jól konzerváltak egy X és Y kromoszómát jelentenek, az Y kromoszóma összeszerelése különösen problémás lehet az említettek miatt.

A diploid és poliploid genomokban egyaránt megtalálható heterozigotitás összeszerelésben okozott problémáját korábban a heterozigotitás redukciójával próbálták csökkenteni. Ennek módszere lehetett beltenyészített vonalak alkalmazása – pl. házinyúl esetében a Thorbecke beltenyészített vonal az OryCun2 genom összeszerelésénél (GenBank: AAGW00000000.2) -, vagy doubled haploid (DH) vonalak alkalmazása (Berthelot *et al.*, 2014; Linsmith *et al.*, 2019). Ezek a módszerek azonban limitáltak, a vizsgált organizmusok természetéből adódóan. Ilyen okok lehetnek a DH vonalak előállításának limitáltsága mindössze néhány fajra, a vadon élő, sokszor

veszélyeztetett, ritka fajok esetében pedig a tenyésztés kizártsága. A heterozigotitásból adódó problémák poliploid genomok esetében tovább súlyosbodnak (Kyriakidou *et al.*, 2018).

Az egyes gének és strukturális variációk egyedenként eltérő jelenléte vagy hiánya további problémát jelent. Extrém esetben ez a gének akár 30%-át is érintheti (Gerdol *et al.*, 2020).

3.5.2. A referenciagenom összeszerelés folyamata

Az első lépés mindig a szekvenálási adatok minőségi szűrése, ezt követi az egymással rések nélkül átfedő szekvenciák összeszerelés, ezek az úgynevezett contigok. Ehhez egymással átfedő leolvasások szükségesek. Minél hosszabb és jobb minőségű leolvasások állnak rendelkezésre, és minél nagyobb lefedettséggel, annál jobb eredménnyel jár ez a lépés.

Attól függően, hogy a további munkához meg akarjuk-e tartani mindkét haplotípust, vagy a genom pszeudo-haploid reprezentációját szeretnénk összeállítani, a contigokat két szettre szeparáljuk. Ezt ma már automatizáltan képesek programok elvégezni, ilyen szoftver például az eredetileg kifejezetten PacBio HiFi technológiához fejlesztett – ma már ONT típusú leolvasásokkal is jól használható – Hifiasm (Cheng *et al.*, 2021). A Hifiasm az összeszerelési lépést követően automatikus haplotípus szeparációt végez. A továbbiakban elsődlegesként kezelt genom verziót, a kettő teljessége alapján választjuk ki. K-mer alapú evaluációs módszerekkel a k-merek „k” hosszúságú, rövid szekvenciákat jelentenek –, mint amilyen a merqury is (Rhie *et al.*, 2020b), könnyen vizsgálható, hogy a teljes k-mer készlet mekkora hányada található meg az egyes verziókban. Érdemes ekkor hasonló minőségi ellenőrzési lépéseket is elvégezni, mint majd a végső összeszerelésen is. Így nyomon követhető nem csak az esetleges hibák megjelenése, hanem a teljes genom minőségének és teljességének lépésenkénti változásai, fejlődése is.

Az első lépést követően hosszabb távú optikai (BioNano) vagy fizikai közelségi (proximity) és térbeli konformáció információ (Hi-C: Belaghzal *et al.*, 2017; Lieberman-Aiden *et al.*, 2009) alapján történik a scaffolding. Az említett technológiák által szolgáltatott információk alapján a kromoszómán belül egymáshoz közel elhelyezkedő szakaszok résekkel kiegészített egymáshoz kapcsolása. Ez az információ egyben felhasználható a contig összeszerelési lépésben történt téves összeszerelések korrekciójára is.

Korábban bevett gyakorlat volt a „polishing”, azaz az összeszerelt genom nagy pontosságú rövid leolvasásokkal történő javítása a hosszú leolvasásokra alapuló összeszerelés után. Ez ma már közel sem annyira szükséges, a PacBio és az ONT technológia közelmúltban jelentősen megnövekedett pontossága miatt (Sereika *et al.*, 2022; Wenger, A. M. *et al.*, 2019). A scaffoldingot követő további korrekciós lépés a Hi-C-re alapuló vizuális térképek alapján történő kézi kuráció. A kézi kuráció elterjedt eszközei a HiGlass (Kerpedjiev *et al.*, 2018) és a PretextView (<https://github.com/sanger-tol/PretextView>). A PretextView előnye az interaktivitás, míg a HiGlassé a nagy felbontás. A PretextView lehetőséget ad a szekvenciák kézi átrendezésére, amelyet a szoftver eltárol, és agp fájl formátumban képes szöveges kimeneti fájl generálni belőle, amely könnyen fasta formátumú szekvencia fájl alakítható. A HiGlass ezzel szemben kizárólagosan vizualizációra alkalmas eszköz. A kuráció során általában problémás szakaszok közeli vizsgálatára, illetve telomer szekvenciák azonosítására használjuk. Bár a PretextView is képes telomer szekvenciák felismerésére – és ez számos esetben elégséges is a kuráció szempontjából –, kevésbé megbízható és pontos, mint a HiGlass.

A „polishing” és a Hi-C alapú scaffolding és korrekció sikerének egyaránt feltétele, hogy a szekvenálás ugyanabból az egyedből történjen, amelyből a hosszú leolvasásos adatok is

származnak. Más egyed felhasználása abban az esetben működhet jól, ha a fajban alacsony a heterozigotitás, de az eredmény minősége nagy valószínűséggel el fog maradni az azonos egyed használatával kapottól.

A cél organizmusban megtalálható organellek genomjai általában külön kerülnek összeszerelésre, több okból. Méretük miatt a teljes genom szekvenálásakor „elvesznek” a nagy mennyiségű, hosszú nukleáris szekvencia között, vagy a nukleáristól eltérő genomösszetételük miatt kontaminációként kiszűrésre kerülhetnek. Ezért ugyanabból az egyedből érdemes külön szekvenálást futtatni, izolált vagy dúsított organellek, illetve organelle DNS felhasználásával.

3.5.3. Referenciák minőségi értékelése

A genom összeszerelése előtt, közben és után is fontos figyelemmel kísérni annak minőségét. A szekvenciák taxonok közti variabilitása miatt nehéz egységes feltételeket felállítani, de érdemes támaszkodni az egyes konzorciumok által felállított alapelvekre, mint például állati genomok esetén a VGP konzorciumé (Rhie *et al.*, 2020a).

Az előzetes minőségi kontroll részét képezi a szekvenálás általános minőségi evaluációja: a leolvasások hosszának, a bázis hívás megbízhatóságának, és a szekvenciákban maradt esetleges adapter szekvenciák detektálása. A szekvenálás minőségéről átfogó képet ad a FastQC (Andrews, 2010). Több futás esetén a MultiQC-vel (Ewels *et al.*, 2016) egyszerűbben összehasonlíthatóvá tehetőek az eredmények. Az ellenőrzést követően célszerű egy „trimming” lépést beiktatni a gyenge minőségű és maradék adapter-szekvenciák eltávolítására. Ezt a lépés akkor is érdemes elvégezni, ha a szekvenciák minősége alapján erre nem feltétlenül lenne szükség, mivel egy-két nem megfelelő minőségű szekvencia is befolyással lehet a végső genom minőségére. A szekvenálási technológiától és személyes preferenciától függően gyakran használt szoftverek például a Trimmomatic (Bolger *et al.*, 2014), amely egy gyors, megbízható szoftver rövid leolvasásokra, vagy a cutadapt, amely hosszú leolvasásokhoz is alkalmas (Martin, M., 2011).

Az ilyen módon ellenőrzött leolvasásokon végzett k-mer analízissel – előre megbecsülhetők egyes összeszerelés során várt értékek, egyúttal előre rámutathat potenciális problémákra. Ilyenek lehetnek a nem várt ploiditásra utaló jelek, nagy mennyiségű repetitív szekvencia jelenléte, vagy a vártnál magasabb hibaérték, amely nagy valószínűséggel az előzetes minőségi szűrés hiányosságára utal.

A k-mer analízis képet ad a genomszekvenálás minőségéről, komplexitásáról és teljességéről is (Simpson, 2014), beleértve a várható heterozigotitást és lefedettséget is. Népszerű, könnyen használható eszköz a k-mer analízishez a GenomeScope (Vurture *et al.*, 2017). Nehezen értékelhető k-mer spektrum esetén célszerű az analízist módosított k-mer hosszúsággal megismételni. Amennyiben nem kapunk a vártak körülbelül megfelelő k-mer spektrumot a szokásos beállításokkal, a mintavétel – jelentős mennyiségű kontamináció –, a szekvenálás, vagy a megelőző minőségi ellenőrzés problémáira utal. Jól beállított k-mer hosszúság esetén diploid genomnál tipikusan a homozigóta szakaszokra jellemző k-merek kétszer, a heterozigóta szakaszokra jellemzőek egyszer fordulnak elő. Nagy mennyiségű repetitív szakasz esetében a ploiditás többszörös értékeinél is megjelennek alacsonyabb csúcsok. A k-mer frekvenciát az összes k-mer mennyisével szemben plottolva vizuális képet kaphatunk a genom várt heterozigotitásáról. Poliploid genomok esetén a spektrum szintén tipikus, de valamivel nehezebben értelmezhető és kevésbé egyértelmű, mint diploid genomok esetében. K-mer analízisre a fentebb említett,

általánosan használt GenomeScope (Vurture *et al.*, 2017) mellett több más program is elérhető (Manekar és Sathe, 2019; Mapleson *et al.*, 2017), és a megfelelő érték kiválasztásához is létezik már automatizált módszer (Chikhi és Medvedev, 2014).

Az összeszerelt genom mérőszámai közé tartozik a teljes szekvencia hosszúság, a contigok és scaffoldok száma, a rések száma és hosszúsága, az N50, illetve NG50 mérőszámok. Az N50 számok mediánt jelentenek, az összeszerelt szekvencia 50%-a ekkora, vagy ennél hosszabb szakaszokban van. A szám contigokra vagy scaffoldokra is vonatkoztatható. Az N50-nél valamivel pontosabb mérték az NG50, ahol a végső összeszerelt genom helyett a becsült teljes genom hossza vonatkoztatjuk a statisztikát. A statisztikának ez a változata alkalmasabb többféle összeszerelés egymással való összehasonlítására (Gurevich *et al.*, 2013). A gfastats (Formenti *et al.*, 2022) és a QUAST (Gurevich *et al.*, 2013) is gyakran használt program az összefoglaló statisztikák egyszerű, gyors elkészítéséhez.

Az összeszerelt genom egyik fontos mérőszáma annak géntartalma. A jelenlegi sztenderd módszer a BUSCO (Manni *et al.*, 2021; Simão *et al.*, 2015) használata. A BUSCO a „Benchmarking Universal Single Copy Orthologs” kifejezés rövidítése. Ahogy neve is mutatja, ismert egy kópiával rendelkező ortológok azonosításán alapul. Minél közelebbi rendszertani kategóriát tudunk felhasználni a BUSCO keresés futtatásához, annál pontosabb becslést kaphatunk az összeszerelt genom teljességéről és helyességéről. A BUSCO mérőszámai között szerepel a vártak megfelelő egyszeres kópiában megtalált (complete), a részlegesen megtalált (fragmented), a duplikált (duplicated), illetve hiányzó (missing) ortológok száma. A referenciához képest egy azzal még nem rendelkező faj újonnan összeszerelt genomja mindenképpen eltérést fog mutatni, de az eltérés mértéke jó indikátor a minőségre.

A referenciák összeszerelésében fontos az esetleges kontamináció szűrése, az összeszerelést megelőzően, utólagos ellenőrzési lépésként is. A kontaminációszűrés eszközei a fentebb leírt K-mer analízis mellett a szekvenciatalajdonságok – például GC-arány, átlagos contig hosszúság – elemzésén alapuló BlobToolKit (Challis *et al.*, 2020), vagy az NCBI által eredetileg belső használatra fejlesztett, de publikusan is elérhető FCS (Foreign Contamination Screen) bioinformatikai eszközcsoomag (<https://github.com/ncbi/fcs>).

3.5.4. A genomösszeszerelés és a lineáris referenciagenomok kihívásai és limitációi

Az egyik legfontosabb limitáció, amelynek számos faj esetében nagy jelentősége van, hogy kis méretű organizmusokból nehéz kellő mennyiségű DNS-t izolálni. Egy egyed esetében a mintát akár teljes egészében szükséges lehet felhasználni egy szekvenáláshoz, több minta bevonása és poolozása pedig a látszólagos heterozigotizást növeli. Veszélyeztetett fajok esetében ezen kívül kerülni kell az invazív mintavételt. Ma egyre kisebb mennyiségű inputtal lehetséges a hatékony genomszekvenálás. A PacBio Low DNA input módszerével például egyetlen, kis méretű szúnyogból (*Anopheles coluzzii*) 100 ng izolált DNS felhasználásával is sikerült jó minőségű referenciát összeszerelni (Kingan *et al.*, 2019), ugyanakkor ennek még mindig feltétele a nagy molekulású DNS izolálása, amelyhez a minta jó állapota szükséges. Szimbiotikus organizmusok esetén pedig egy minta akár több teljes genomot is tartalmazhat, amelyek elválasztva, külön kerülnek összeszerelésre (Challis *et al.*, 2020).

A klasszikus referenciagenommal szemben, amely egy egyed alapján készülve annak teljes genomját és minden változatát tartalmazza, ezzel reprezentálva a fajt, a pangenom gráfok

igyekeznek a faj teljes diverzitását felölelni (Paten *et al.*, 2017). Ez különböző mértékben lehetséges, elsősorban a vizsgált faj, illetve taxon befolyásolja. Az emlős genomok aránylag kisebb diverzitása miatt a pángenom gráfok közelebb juthatnak a teljes genom leírásához, míg más taxonokban nagy mennyiségű gén jelenléte vagy hiánya akár a vizsgált egyeden is múlhat (Gerdol *et al.*, 2020). A pángenom reprezentáció ezt a problémát orvosolhatja, ugyanakkor minél kiterjedtebb a gráf, annál több számítási kapacitást igényel a vele történő munka (Kaye és Wasserman, 2021). A humán pángenom referencia például 47, teljes egészében haplotipizált, diverz diploid humán genom felhasználásával készült (Liao *et al.*, 2023) – más fajok esetében azok genetikai diverzitásától függően ennél sokkal több egyed lehet szükséges a megfelelő reprezentációhoz.

3.5.5. A referencia genomok hatása az analízisre

A referenciagenom-választás hatását a közelmúltban egyre több tanulmány vizsgálta, igazolva a referencia befolyását az analízisre, és a megfelelő minőségű, a cél fajhoz közel álló referenciák szükségességét.

A választott referencia, illetve referenciák hatását a leolvasások illesztésére és a további elemzésre vizsgálták mikrobiális genomok esetében Valiente-Mullor és munkatársai (2021). Van Bemmelen és munkatársai (2025) szintén mikrobiális genomokat vizsgálva mutatta meg, hogy a megfelelő referencia választása javítja a taxonómiai klasszifikációt. Humán genomikai vizsgálatok alapján pedig kiderül, hogy a választott referencia verzióknak már fajon belül is jelentős hatása lehet. A humán GRCh37 és GRCh38 genom verziók közötti eltérések a hívott SNP-k és indelek körülbelül 1,5%-ának eltérést eredményeznek. A referenciából adódó eltérések közül több is betegségekkel asszociált (Li, H. *et al.*, 2021). A referenciagenom az RNS-szekvenálási eredmények értelmezésére is hatással van: az illesztési eltérések következtében eltérő expressziós értékeket kaphatunk, amely fehérjekódoló és ritka betegség géneket is érinthet (Ungar *et al.*, 2024).

Nem modell szervezetek esetében a fajok túlnyomó többségéhez nem létezik fajspecifikus referenciagenom. Egyes fajok esetében draft genomok léteznek – olyan genomok, amelyek bár reprezentálják a faj genomját, nincsenek kromoszóma szinten összeszerelve –, de ezzel együtt is csak kevés faj teljes genomjáról áll rendelkezésünkre információ. Általában a fajok között nagyobb a genetikai variabilitás, mint fajon belül. Ez befolyásolhatja a leolvasások illesztését, a variációk hívását, és az abból levont következtetéseket is. Faj- és fajtamegőrzési szempontból is fontos az olyan referenciagenom, ami alapján megbízható eredményeket kapunk.

Nevado és munkatársai (Nevado *et al.*, 2014) szimulált referenciákat és adatokat alkalmazva vizsgálták a referenciagenomok hatását, míg más kutatások rokon fajok draft- és referenciagenomjain végeztek összehasonlításokat. Vizsgálták például a cél fajból létező draft genom és egy közeli rokon faj referenciájának használatából adódó különbségeket (Shafer *et al.*, 2017), vagy a vad farkas, és a háziásított kutya referenciája befolyását vad kutyafélék populációgenetikai elemzésére (Gopalakrishnan *et al.*, 2017). Bohling (2020) pedig lazac (Salmonidae) fajokban vizsgálta RADseq adatok felhasználásával 6 különböző Salmonidae faj referenciájának befolyását az egyes adatokon végzett elemzés eredményeire. Szarvasmarha fajtákban végzett vizsgálatokban a referencia befolyása nem adódott jelentősnek, de szintén kimutatható volt (Lloret-Villas *et al.*, 2021).

A referenciagenomok hatását vizsgáló publikációk általános következtetése, hogy a vizsgált fajtól távolabb eső referencia a teljes leolvasáskészletből kisebb arányú és gyengébb minőségű leolvasás

illeszkedéshez vezet. A hívott variációk mennyisége és minősége is változónak adódott. Ezutóbbi a választott hívó algoritmuson és a referencia távolságán is múlik. A populációk között mért differenciációs és fixációs szintben, és a filogenetikai klaszterezésben is minden esetben különbség mutatkozott.

A populációgenetikában elterjedten használt RADseq adatok különösen referenciaérzékenyek lehetnek, a restrikciós site-okhoz való kötöttség miatt: az könnyen egyes lókuszok „ elvesztéséhez” vezethet a fajok között (Arnold *et al.*, 2013; Davey *et al.*, 2013). A legnagyobb hatás olyan lókuszokon várható, amelyek eltérő allélra fixáltak (Lloret-Villas *et al.*, 2021; Van Bemmelen *et al.*, 2025).

Összességében a genotipizálás minősége gyakorolja a legnagyobb hatást az elemzés végeredményére, az pedig a referencia és a bioinformatikai munkafolyamat együttes következménye. Az illesztés jóságának befolyása különösen nagy, mivel minden további folyamatra hatással van.

4. ANYAG ÉS MÓDSZER

4.1. Teljes genom szekvenálás alapú populációgenetikai vizsgálatok vadon élő és házinyúl fajtákban

4.1.1. Mintagyűjtés és DNS izolálás

Az üregi nyúl (*Oryctolagus cuniculus*) végzett populációgenetikai vizsgálatához 14 egyedet használtunk fel, két magyar üregi nyúl populációból. Ezek a Tiszaszentimre környékén vadon, és a Budapesti Állat- és Növénykertben fogságban élő populációk, amelyekből 6, illetve 8 egyed került szekvenálásra. Házinyúl mintákat a következő fajtákból szekvenáltunk: Rex (5 egyed), Tübingiai (1 egyed), Kaliforniai (3 egyed), Hycote (10 egyed), Zika (6 egyed), És Új-Zélandi fehér (3 egyed) ([M2. Melléklet](#)).

A minták vér- és fülszövetből származnak, kereskedelmi nyúltelepekről. A minták tárolása -20 °C-on történt a DNS izolálásáig.

Mintánként körülbelül 20 mg fülszövetből történt DNS kivonás, MagCore Genomic DNA Tissue Kittel (RBC BioScience Corporation, Taiwan). A mintákat 400 µl GT-puffer és 20 µl (10 mg/ml) proteináz-K enzim elegyében, thermomixerben 55 °C-on 90 percig homogenizáltuk, majd filteroszlopon történő szűrést követően, 5 percig maximumon centrifugáltuk. A DNS izoláláshoz a MagCore HF16Plus Automated Nucleic Acid Extractor (RBC Bioscience Corporation) robotot használtuk. Az elúciós térfogat 200 µl volt.

Vérminták esetében 400 µl térfogatú mintából a the MagCore Genomic DNA Whole Blood Kit (RBC Bioscience Corp., Taiwan) használatával, a továbbiakban azonos módszerrel történt az izolálás, az elúciós térfogat ebben az esetben is 200 µl volt.

DNS koncentrációt NanoDrop ND-1000 (Thermo-Fisher Scientific, USA) készülékkel ellenőriztük. A DNS izolálást és minőségellenőrzést a MATE-GBI Alkalmazott Vad- és Haszonállat Genomikai Csoportja végezte.

4.1.2. Szekvenálás

Összesen 42 minta került teljes genom szekvenálásra a projekthez. A könyvtárkészítést és a szekvenálást a Novogene végezte, sztenderd teljes genom szekvenálási (WGS) protokollnak megfelelően. Az Illumina NovaSeq 6000 készüléken páros (paired-end), 150 nukleotid hosszúságú leolvasásokkal készült a szekvenálás, az inzert méret 300 nukleotid volt, az átlagos lefedettség 35.4 (±6.37)-nek adódott.

4.1.3. Publikus adatok gyűjtése

A munkám során felhasznált publikus adatok az NCBI SRA-ról származnak. Az alkalmas mintákat az összes, 2022.05.30-án elérhető üregi nyúl genomi szekvenciából választottam ki. A kiválasztási feltételek között szerepelt, hogy a minta teljes genom szekvenálásból származzon, rövid leolvasásos, lehetőleg Illumina szekvenálás legyen, valamint rendelkezésre álljon megfelelő teljességű háttérinformáció (metadata). Ez utóbbi esetében különösen fontos volt, hogy ismert legyen a házinyúl fajta, vagy a vadon élő üregi nyúl minta földrajzi eredete.

Külön figyelmet fordítottam arra, hogy egyértelműen eldönthető legyen, hogy poolozott, azaz egy szekvenálásban több egyedből származó szekvenciákat tartalmazó, vagy egyedi szekvenálásból származnak az adatok. A poolozottság nem számított általános kizáró oknak, de egyes tesztekre ezek a minták nem alkalmasak.

Minden mintát, amely nem teljesítette az összes szükséges feltételt, automatikusan kizártam. A publikus minták átlagos lefedettsége 9.5-nek adódott. A legalacsonyabb lefedettség 3,9-szeres, míg a legmagasabb 14.3-szoros volt.

Összesen 56 publikus mintát választottam ki és használtam fel, amelyek részletes listáját az [M2. melléklet](#) tartalmazza.

4.1.4. Minőségi kontroll és genetikai variációk azonosítása

A minőségi ellenőrzés első lépéseként a FastQC v.0.11.7 (Andrews, 2010) programmal ellenőriztem azt, majd a Trimmomatic v.0.36 (Bolger *et al.*, 2014) szoftverrel eltávolítottam az adapter szekvenciákat és rossz minőségű leolvasásokat. Az alkalmazott paraméterek a következők voltak: csúszóablak mérete (sliding window size) 30, lépés méret (step size) 4 nukleotid, legkisebb megtartott leolvasás hosszúság (minimum read length) 50, illetve a leolvasások végén előforduló gyengébb minőségű szakaszok esetében a (leading and trailing cutoff) 28-as quality score alatti nukleotid szekvenciák kerültek levágásra. A minőségi szűrésen megfelelt adatokat az *OryCun2.0* referenciagenomra (NCBI RefSeq GCF_000003625.3, GenBank GCA_000003625.1) (Carneiro *et al.*, 2014b) illesztettük a bwa mem v.0.7.17 (H. Li és Durbin, 2011) programmal, alapbeállításokkal.

Az illesztést követően a GATK ToolKit (v.4.1.8) (Auwera és O'Connor, 2020) által ajánlott jó gyakorlatot követtük (DePristo *et al.*, 2011), saját adatainkhoz igazítva azt. Az illesztést követően a GATK MarkDuplicates programjával eltávolítottuk azokat a leolvasás párokat, amelyek nagy valószínűséggel ugyanannak a DNS fragmensnek a duplikációjából (például PCR duplikátum) származott. A program alapvetően megjelöli a vélt duplikációkat, a megfelelő opciókkal pedig egy kivétellel el is távolítja azokat: `--REMOVE_DUPLICATES` a könyvtárkészítésből származó duplikációk, `--REMOVE_SEQUENCING_DUPLICATES` pedig a szekvenálási artefaktok eltávolításához. A duplikációk kiszűrését követően a quality score recalibrációt végeztünk a Base Score Recalibrator (BQSR) programmal. A BQSR két részből áll, a BaseRecalibrator-ból, amely az új értékeket kalkulálja, és az ApplyBQSR-ből, amely ezekből elkészíti az új bam fájlt, a megfelelő minőségi értékekkel.

A szűrési és recalibrációs lépések után következett maguknak a variációknak a meghatározása. A variáció hívást (variant calling) a GATK HaplotypeCaller programjával végeztem. A variációk hívása ekkor mintánként történik, ennek az eredményét további genotipizálási és szűrési lépésekkel konszolidáltam. Az első lépés az egy-egy mintát tartalmazó fájlok kombinálása a CombineGVCFs programmal, majd ezt követi azok együttes genotipizálása (joint genotyping) a GenotypeGVCFs-szel. Ezeknél a lépéseknél a GATK által ajánlott alapbeállítások jól működtek.

A változatok minőségi szűréséhez a „kemény” szűrési módszert választottuk. Az elsősorban ajánlott módszer a gépi tanulás alapú Variant Quality Score Recalibration (VQSR) algoritmus lenne, ennek megbízható működéséhez azonban nagy mennyiségű, kiváló megbízhatóságú ismert polimorfizmust kéne felhasználni tanuló szettként. Mivel ehhez megfelelő adatbázis ismert nyúl SNP-kből nem áll rendelkezésre, a VariantFiltration szoftver merev szűrési módszerét választottuk. A szűréshez általam beállított paraméterek azonosak voltak a GATK által ajánlottakkal: `-filter "QD < 2.0" --filter-name "QD2" -filter "QUAL < 30.0" --filter-name "QUAL30" -filter "SOR > 3.0" --filter-name "SOR3" -filter "FS > 60.0" --filter-name "FS60" -filter "MQ < 40.0" --filter-name "MQ40" -filter "MQRankSum < -12.5" --filter-name "MQRankSum-12.5" -filter "ReadPosRankSum < -8.0" --filter-name "ReadPosRankSum-8" SNP-k és -filter "QD < 2.0" --filter-name "QD2" -`

`filter "QUAL < 30.0" --filter-name "QUAL30" -filter "FS > 200.0" -
-filter-name "FS200" -filter "ReadPosRankSum < -20.0" --filter-name
"ReadPosRankSum-20"` indelek szűréséhez. Mivel az indelek és SNP-k eltérő paraméterekkel szűrendők, előzetesen a genotipizált vcf fájlból a SelectVariants programmal kiválasztottam az SNP-ket egy fájlba, egy másikba az indelek és kombinált lókuszok összességét. A kombinált lókuszok esetében egy genomi koordinátán SNP és indel is hívásra került. A szűrés után a két fájlt ismét egyesítettem, a MergeVcfs programmal.

A variációk azonosítását és minőségi szűrését követően allél gyakoriságra – minimum 0.001 a minor allél esetében – allélszámra – minimum 2, maximum 2, azaz biallélikus variációk –, és kapcsoltági egyensúlyhiányra is szűrtem az adatokat. Ezekhez a szűrés lépésekhez a vcftools 0.1.17 (Danecek *et al.*, 2011) és a PLINK 1.9 (Chang *et al.*, 2015) szoftvert használtam.

A teljes genom lefedettsége lehetőséget adott a szigorú szűrésre. Vizsgálataim során kizárólag biallélikus variációkat használtam, alkalmazkodva egyes szoftverek limitációihoz. Ezek között SNP-k és indelek is szerepeltek. Hiányzó genotípusokat egyáltalán nem engedtem meg, azokat a genomi pozíciókat, ahol akár egy minta genotípusa is hiányzott, kizártam. Csak az elsődleges, kromoszómákon elhelyezett szekvenciákat („primary”) használtam fel.

4.1.5. Populáció-struktúra és beltenyésztettség

A főkomponens-elemzéshez az eigenvector értékek kalkulációját a PLINK 1.9 (Chang *et al.*, 2015) segítségével végeztem. A maximális valószínűségeen alapuló filogenetikai fát a biallélikus variációk alapján az SNPhylo pipeline (Lee *et al.*, 2014) használatával készítettem. Az SNPhylo pipeline egy vcf fájlból kiindulva, az SNP-kből generált szekvenciák egymáshoz illesztésével kalkulálja a filogenetikai fát.

A házinyúl és az üregi nyúl állomány között előforduló hibridizáció detektálására először az ADMIXTURE v1.3 (Alexander *et al.*, 2009) programot használtam. Első lépésként cross-validation errort kalkuláltam a lehetséges felmenő populációk számára (K), 1 és 12 között. A legalacsonyabb hibát a K=5 és K=7 értékek adták. Az ADMIXTURE-t követően a Treemix v.1.13 (Pickrell és Pritchard, 2012) szoftverrel is megvizsgáltam a populációk közötti lehetséges hibridizációt. Ehhez a vizsgálathoz csak az általunk szekvenált mintákat használtam fel, és outgroupként egy Ibériai mintát az üregi nyúl *Oryctolagus cuniculus algirus* alfajából. A két magyar üregi nyúl populációból származó mintákat az analízis szempontjából két külön populációként kezeltem. Ötszáz bootstrapping lépéssel futtattam a programot. A fentiekén kívül D (ABBA-BABA) és f4 teszteket is futtattam, az AdmixTools (<https://github.com/DReichLab/AdmixTools>) és admixr (Petr *et al.*, 2019), <https://bodkan.net/admixr/>) R csomagok felhasználásával.

A beltenyésztettség mértékét a magyar üregi nyúl populációkban a genomban található homozigóta szakaszok hosszának és mennyiségének meghatározásával vizsgáltam. Ezek az úgynevezett „runs of homozygosity” (ROH) szakaszok. Ebből a vizsgálatból a poolozott mintákat kizártam, mivel több egyed együttes szekvenálása és genotipizálása során a heterozigotitás általában magasabbnak tűnik, mint amilyen valójában. A ROH szakaszok detektálására a PLINK 1.9-et (Chang *et al.*, 2015) használtam, amelyet követően ezeket a szakaszokat R és bash scriptek segítségével tovább kategorizáltam, számszerűsítettem és vizualizáltam. A minimálisan figyelembe vett ROH szakasz hosszúságát 100 kbp-ben határoztam meg, amely felett további kategóriákat határoztam meg, legfelső kategóriaként az 1 Mbp-nél hosszabb összes szakasszal.

4.1.6. Szelektált genomi régiók detektálása és jellemzése házinyúl fajtákban

A szelekció alatt álló genomi régiókat három módszer kombinációjával határoztam meg. Ezek a composite likelihood ratio (CLR), nukleotid diverzitás (π) és a fixációs index (Fst).

A CLR tesztet a SweepFinder2 (DeGiorgio *et al.*, 2016) szoftver használatával végeztem. A SweepFinder2 először egy empirikus háttérfrekvencia-spektrumot kalkulál a jelenlévő allélekből a tényleges teszt futtatása előtt, ebből számítja ki a nullhipotézis valószínűségét. A kalkulációhoz körülbelül 10 kbp-s ablakokat használtam. A körülbelüliség oka, hogy a CLR által tesztelt pontok pontos elhelyezkedése a variációk pontos koordinátáin is múlik. A szoftver automatikusan választja ki a megfelelő pontot a felhasználó által megadott genomi távolságok alapján.

Az Fst értékek számításához a VCFTools v0.1.17-ben (Danecek *et al.*, 2011) implementált Weir és Cockerham-féle módszert (Weir és Cockerham, 1984) használtam (`--weir-fst-pop`, `--fst-window-size`, `fst-window-step` opciók). A VCFTools segítségével kalkuláltam a nukleotid diverzitási értékeket is (`--window-pi`, `--window-pi-step` opciók). Mindkét esetben csúszóablakos megközelítést használtam, 10 kb hosszú ablakokkal és 5 kb-s lépésekkel.

Az ablakok és lépések méretét úgy választottam meg, hogy a felbontás lehetőség szerint konzisztens maradjon a három módszer között és kellő mennyiségű mérési pont legyen, de ez ne járjon túlzott számítógépes kapacitás igény növekedéssel.

Mindhárom teszt 99. percentilisét, azaz az Fst és a CLR esetében a legmagasabb, míg a nukleotid diverzitás esetében a legalacsonyabb értékeket határoztam meg, mint potenciálisan szelektált régiókat. A jelölt régiók átfedő szakaszait a BEDTools (Quinlan és Hall, 2010) intersect programmal kerestem meg. Az ilyen módon meghatározott régiók közül a szomszédosakat összeolvasztva, egy régióként kezeltem a továbbiakban. Szomszédosnak tekintettem azokat a régiókat, amelyeknek egymáshoz legközelebbi koordinátái egymáshoz képest 10 kbp-n belül adódta. Az így meghatározott régiók géntartalmát az Ensembl (release 104) (Howe, K. L. *et al.*, 2021) annotációja alapján határoztam meg. A szelektáltak adódott géneken génontológiai (GO) tesztet is végeztünk. Ehhez legalább két módszerrel szelektált régióba eső, vagy azzal átfedő géneket vettem figyelembe. Mivel az elérhető GO adatbázisokban nem voltak közvetlenül megtalálhatóak az üregi nyúl gének, ezért a humán és egér ortológokat használtam a tesztekhez. Az ortológok meghatározásához az Ensembl BioMart (release 109) (Kinsella *et al.*, 2011) adatbázisát használtam. A Fisher exact tesztet a Panther GO-Slim adatbázis biológiai folyamat („biological process”), molekuláris funkció („molecular function”) és útvonal („pathway”) kategóriáiban futtattam.

A nyúl genomon található lehetséges transzkripció faktor kötőhelyeket (TFBS) a Liftover (Hinrichs, 2006) program segítségével határoztam meg. A Liftoverrel a ChipSummitDB humán TFBS-adatbázis (Czipa *et al.*, 2020) alapján térképeztem a nyúl genomra a kötőhelyeket. A potenciális TFBS koordinátákat összevettem a potenciálisan szelektált régiókkal, az egyes kötőhely típusok dúsulását pedig Fisher exact teszttel vizsgáltam.

4.2. Kromoszóma szintű *Lepus timidus* referenciagenom összeszerelése

4.2.1. Mintagyűjtés és fajazonosság megerősítése

A referenciagenom összeszerelésének alapjául szolgáló fiatal hím havasi nyúl (*Lepus timidus*) egyedét 2018 októberében lőtték ki, Finnországban, Ilomantsi területén, a vadászati időszakban. A vadászati terület emberlakta területektől távol eső, tajga típusú erdő. A Finnországban élő egyetlen másik, a *Lepus* nemzetségbe tartozó faj, a mezei nyúl (*Lepus europaeus*) emberalkotta, vagy ember által erősen befolyásolt környezetben fordul csak elő az országban, így a mintavételi

terület megválasztása csökkentette annak az esélyét, hogy hibrid egyed került volna mintázásra. (Levänen *et al.*, 2018b). A fajazonosság a kilövést követően a havasi nyulat a mezei nyúltól elkülönítő morfológiai jegyek alapján került megállapításra. Szekvenálást követően kódoló szekvenciák és mitokondriális haplotípus vizsgálatával megerősítésre került az egyed nem-hibrid havasi nyúl státusza. A nukleáris markerek vizsgálatát részletesen Gaertner *et al.* (2023) munkája írja le, amely szerint az LT1 sejtvonal 9056 SNP és több, mint félmillió transzkript elemzése alapján nem mutatta hibridizáció jeleit.

Az egyedből immortalizált fibroblaszt sejtvonal került létrehozásra (LT1; Gaertner *et al.*, 2023) majd ebből került izolálásra DNS a szekvenálásokhoz. A sejtvonal elérhető a ZFMK-TIS-69753 voucher számon az Alexander König Múzeumnál (ZFMK) (Bonn, Németország).

4.2.2. DNS izolálás és szekvenálás

A referenciagenom elkészítéséhez három szekvenálást végeztünk.

A PacBio HiFi szekvenáláshoz szükséges nagy molekulású DNS izolálása fenol-kloroform extrakciós módszerrel történt, Sambrook és Russel (Sambrook & Russell, 2006) módszerét követve. A DNS Qubit (Thermo-Fisher Scientific, USA) készülék segítségével lett kvantitálva, majd a fragmenthosszúság a Fragment Analyzer (Agilent Technologies, USA) készülékkel, a Large Fragment Analysis Kit (Agilent Technologies, USA) segítségével került meghatározásra. A könyvtárkészítést és szekvenálást a Helsinki Egyetem Biotechnológia Intézetének „DNA Sequencing and Genomics Laboratory” csoportja szolgáltatásként végezte (Fekete *et al.*, 2025a).

A Hi-C szekvenálási könyvtár Belaghzal és munkatársai (2017) protokollja alapján készült, négy módosítással. A protokoll 3 ismétlésben lett végrehajtva; a méretfrakcionálás a NEBNext Ultra II FS DNA modullal történt; a könyvtár „NEBNext Ultra II library preparation kit for Illumina” (New England Biolabs, USA) kittel készült; a PCR reakciók triplikátumban, egyenként 6 PCR ciklussal lettek végrehajtva. A PCR termékek Ampure XP (Beckman Coulter Life Sciences, USA) gyöngyökkel kerültek tisztításra. A végső, tiszta könyvtárakat Qubit készülék segítségével kvantifikáltuk, majd agaróz gélelektroforézissel megerősítettük a fragment mértéket. A szekvenálás Illumina NovaSeq 6000 készüléken történt, SP flowcell használatával. A szekvenálás páros leolvasásos (paired-end) volt, a readok hosszúsága 2 x 150 nukleotid.

Míg a teljes genom szekvenáláshoz a biobanki mintából létrehozott fibroblaszt sejttenyészetből izoláltunk DNS-t, a sejtvonal mitokondriális DNS-ének izolálása a biobankban tárolt fagyasztott fülszövetből történt, a peqGOLD blood & tissue DNA mini kittel (VWR), a gyártó protokolljának megfelelően. (Tapanainen *et al.*, 2024). A szekvenálás PCR amplifikációt követően, Illumina MiniSeq készüléken történt, a Kelet-Finnországi Genomikai Központban (Itä-Suomen Genomikeskus (ISGK), Kuopio, Finnország). A mitokondriális szekvencia külön is elérhető a GenBank #OR915850 azonosító alatt.

4.2.3. Minőségi kontroll

A PacBio HiFi és Illumina Hi-C szekvenálási adatok minőségét egyaránt a FastQC v.0.11.9 (Andrews, 2010) programmal ellenőriztem, majd a cutadapt v.4.6-tal (Martin, 2011) eltávolítottam a maradék rossz minőségű és adapter szekvenciákat. Az összeszerelés előtt ellenőriztem a PacBio adatok k-mer profilját a Meryl v1.3 (<https://github.com/marbl/meryl>) és a GenomeScope v2.0 (Vurture *et al.*, 2017; Ranallo-Benavidez *et al.*, 2020) programokkal, majd az eredményeknek megfelelően állítottam be az összeszerelési paramétereket.

4.2.4. Genom összeszerelés

Az összeszerelést a HiFiasm v0.16.1 (Cheng *et al.*, 2021, 2024) programmal kezdtem. A minőségi ellenőrzésen átesett mintákon a következő opciókkal futtattam le a programot: `-m 100000000 -p 100000 -hom-cov 20`, valamint a Hi-C szekvenálási adatok integrációjához a `-hic1 -hic2`. A HiFiasm által generált nyers genom tulajdonságait a gfastats v1.3.6 (Formenti *et al.*, 2022) és a mercury v1.3 (Rhie *et al.*, 2020b) segítségével vizsgáltam, és a két összeszerelt genom közül a statisztikák alapján teljesebbnek mutatkozót választottam ki „primary”, elsődleges összeszerelésként, míg a másikat „alternate”, azaz alternatív szekvenciaként vittem tovább.

Az összeszerelés közben az alternatív haplotípusok miatt keletkezett duplikált contigokat a `purge_dups v1.2.6` (Guan *et al.*, 2020) segítségével mindkét nyers genomról eltávolítottam. Ezt követően folytattam a scaffoldinggal.

A Hi-C adatok feldolgozásához az ArimaGenomics által kifejlesztett munkafolyamatot követtem. Ennek fő lépései az adatok térképezése az eddig összeszerelt és deduplikált genomra a `bwa-mem2 v2.2.1` (Vasimuddin *et al.*, 2019) programmal, majd a leolvasások válogatása és szűrése a `samtools (v1.18)` (Danecek *et al.*, 2021; Li, H. *et al.*, 2009) és az ArimaGenomics github oldalán elérhető scriptek (https://github.com/ArimaGenomics/mapping_pipeline) segítségével. Az előforduló duplikált leolvasásokat eltávolítottam a GATK toolkit v4.3.0.0 (Auwera és O’Connor, 2020) részét képező `MarkDuplicates` programmal. A tényleges scaffolding lépést a `YaHS v1.2` (Zhou *et al.*, 2023) programmal végeztem. A programot a következő opciókkal futtattam: `-e GATC -l 5000 -no-contig-ec`.

A genom folytonosságát és általános statisztikákat a `QUAST v5.2.0` (Mikheenko *et al.*, 2018), a genom teljességét pedig a `BUSCO v5.5.0` (Manni *et al.*, 2021) program segítségével kalkuláltam. Utóbbit a `glres_odb10` (Lagomorpha) és a `mammalia_odb10` (Mammalia) adatbázisokkal is elvégeztem. A felhasznált adatbázisok elérhetők a <https://busco-data.ezlab.org/v5/data/lineages/> címen.

4.2.5. Telomer- és repetitív szakaszok annotációja

A repetitív szakaszok meghatározását és annotációját a `RepeatModeler v2.0.5` (Flynn *et al.*, 2020) program segítségével végeztem. A `RepeatModeler` által készített repetitív szakasz-könyvtárat használtam a genom maszkolásához a `RepeatMasker v4.1.6` (Smit *et al.*, 2015) programmal, amely fontos előkészítése a genom további minőségi ellenőrzésének és korrekciójának, valamint a génannotációnak is. A génannotációt a genom közzétételét követően az NCBI munkatársai végezték.

A telomer szekvenciákat a `Telomere Identification ToolKittel (tidk v0.2.0)` (Brown *et al.*, 2025) azonosítottam, először „`explore`”, majd „`search`” módban futtatva a programot. Az „`explore`” mode segít meghatározni egy olyan faj potenciális telomer szekvenciáit, amelyben azok még nem ismertek, a „`search`” mode pedig a felhasználó által meghatározott szekvencia alapján keres telomer szakaszokat a genomon. Ez esemben az „`AACCCT`” szekvencia volt.

4.2.6. Kézi kuráció

A scaffoldingot követően a repetitív szakaszokat maszkolt genomon kézi kurációt és további korrekciókat végeztek a Wellcome Sanger Institute (Hinxton, Egyesült Királyság) kurátorai a `PretextView` (<https://github.com/sanger-tol/PretextView/>) és `HiGlass` (Kerpedjiev *et al.*, 2018) szoftverek és a `rapid curation workflow` scriptjei (<https://gitlab.com/wtsi-grit/rapid-curation>)

segítségével. A folyamatot részletesen Howe és munkatársai (Howe, K. *et al.*, 2021) munkája írja le.

4.2.7. Mitokondriális szekvencia összeszerelése

A mitokondriális szekvencia összeszerelését a MitoZ pipeline (Meng *et al.*, 2019) segítségével végeztem. A pipeline-t többféle paraméter beállítással is lefutttam, majd a legjobb eredményt fogadtam el. Az elfogadott szekvencia fő kritériuma a cirkularitás, és az ismert mitokondriális gének teljes jelenléte volt. A végső verzióhoz használt paraméterek a következők voltak: `--clade Chordata --fastq_read_length 150, --requiring_taxa Chordata --genetic_code 2 --kmers_megahit 21 29 39 59 79 99 119 141`.

4.2.8. Összehasonlítás elérhető *Lepus* genom szekvenciákkal

Eredményeinket összevetettük az NCBI-on elérhető korábbi *Lepus timidus* referenciával (NCBI Accession number: GCA_009760805, Marques *et al.*, 2020), valamint az *Lepus europaeus* referenciagenommal (NCBI accession number: GCF_033115175.1, Michell *et al.*, 2024). A genomok egymáshoz illesztését a minimap2 v.2.21 (Li, 2018) szoftverrel végeztem, majd a szekvenciahasonlóságot és -szinténiát a parCoordsDotPlotly.R (<https://github.com/tpoorten/dotPlotly>) scripttel vizualizáltam.

4.3. Lineáris DNS-vezérelt mitokondriális rekombináció kimutatása emlős sejtekben

4.3.1. Felhasznált sejtvonalak

A kísérletekhez két, egy humán, és egy egér eredetű sejtvonalt használtunk fel.

A humán eredetű BH10 sejtvonalt egy HeLa sejteknek egy Kearns-Sayre szindrómás (KSS) páciens fibroblaszt klón sejtvonaltól származó fúziójával létrehozott sejtvonalt, amely így két féle mitokondriális DNS-t tartalmaz, egy vad típust (wt, HeLa sejtekből) és egy deléciót (Δ , KSS sejtekből). A deléció a 8 482.-től a 13 460. nukleotidig terjed, a wt mtDNS koordinátái alapján. A sejtvonalt eredetét és létrehozását részletesen (Sancho *et al.*, 1992) írja le.

A HP202B sejtvonalt osztdó sejtvonallá dedifferenciált elsődleges egér hepatocitákból kerültek létrehozásra (Battersby, 2001; Bayona-Bafaluy *et al.*, 2005). Ez a sejtvonalt New Zealand Black (NZB) és Balb/c egér sejtvonaltok mtDNS-ét tartalmazó heteroplazmikus sejtvonalt (Bacman *et al.*, 2010), amely 90 polimorfizmusban különbözik a mitokondriális genom teljes hosszán.

A mitokondriális rekombináció lehetőségének hatékony vizsgálatához a sejtvonaltok MGME1 knockout változatára volt szükség. Az MGME1 expressziójának megszüntetése, illetve csökkentése a lineáris DNS-fragmentek lebontásától a hibajavítás irányába tolja el a kettős szártörések kezelési mechanizmusát. A BH10 és HP202B sejtvonaltokból CRISPR-Cas9 genom szerkesztési módszerrel készült knockout változat. A guide RNS szekvenciák tervezése egy online elérhető eszközzel (<https://www.benchling.com/>) történt, a target pedig az MGME1 start kodonja volt. A guide RNS-ek létrehozását és a sejtek transzfektálását részletesen Balboa és munkatársai (2015) és Fragkoulis és munkatársai (2024) írják le. Az MGME1 knockout PCR genotipizálással és Western blottal került megerősítésre (Fragkoulis *et al.*, 2024).

4.3.2. DNS izolálás és szekvenálás

A teljes DNS izolálás sejtekből konfluens 6-well plate-ekben, DNS lízis puffer használatával (10 mM Tris pH 7.4, 10 mM EDTA, 150 mM NaCl, 0.4% SDS, 200 μ g/ml proteináz K) történt, ezt követte két fenol-kloroformos extrakció és etanolos kicsapás. Az mtDNS izolálása a BH10 és HP202B sejtvonaltok esetében is hipotóniás pufferrel, majd ezt követő differenciáló

centrifugálással történt. A mitokondriumok elválasztásához 1,5M/1M-os szacharóz gradienst használtunk (50 000 g, 30 perc, 4 °C). A mitokondriumokat mtDNS lízis pufferben lizáltuk (75 mM NaCl, 50 mM EDTA, 1% TritonX-100, 20 mM HEPES pH 7,8, 200 µg/ml Proteinase K). Ezt három fenol-kloroformos extractió lépés, majd izopropanolos kicsapás követte.

A minták ezután -20 °C-on, TE pufferben vagy 20 mM HEPES pH 7,2-ben voltak tárolva. Az mtDNS-t XhoI enzim alkalmazásával linearizáltuk, az emésztés 3 órán át 37 °C-on történt, 1 µl enzim per 5 µg mtDNS alkalmazásával. Ez az enzim a humán mtDNS-t mindkét alkalmazott haplotípus esetében a 465., míg az NZB egér mtDNS-t a 13 561., a BALB/c mtDNS-t pedig a 13 559. nukleotidnál hasítja. Valamennyi mtDNS típus esetében csak egy helyen történik hasítás (Fragkoulis *et al.*, 2024).

Az MGME1 KO sejtek mtDNS-ének szekvenálása PacBio HiFi technológiával (Wenger, A. M. *et al.*, 2019) történt, a *Lepus timidus* genom összeszereléséhez készült teljes genom szekvenáláshoz hasonlóan. A BH10 sejt vonal esetében a leírt módon tisztított mtDNS-ből, míg a HP202B sejtek esetében teljes DNS-ből történt a szekvenálás. A könyvtárkészítést és a szekvenálást ebben az esetben is a Helsinki Egyetem Biotechnológia Intézetének „DNA Sequencing and Genomics Laboratory” csoportja végezte, szolgáltatásként.

A mitokondriális DNS haplocsoportok meghatározása a PhyloTree Build alapú webes elemző programmal történt (<https://dna.jameslick.com/mthap/>) (Van Oven és Kayser, 2009).

4.3.3. Minőségi ellenőrzés és tényleges mitokondriális szekvenciák szelektálása

A minőségi szűrés lépései megegyeztek a humán és az egér sejt vonalakból származó szekvenálási adatok esetében. A leolvasások minőségét először a FastQC v.0.11.9 (Andrews, 2010) szoftverrel ellenőriztem, majd a cutadapt v4.6. (M. Martin, 2011) használatával eltávolítottam bármilyen maradék adapter tartalmat, vagy más szempontból nem megfelelő minőségű szekvenciát. A cutadaptot az **--action trim --error-rate 0.1 --overlap 35 --discard-trimmed** opciókkal futtattam. Ezen kívül a **-b** opcióval megadtam a keresendő PacBio HiFi adapter szekvenciát is.

A szekvenciákat ezután a pbmm2 v.1.13.0. szoftverrel (<https://github.com/PacificBiosciences/pbmm2>), „ccs preset” beállításokkal illeszttem a megfelelő referenciagenomokra. Ez humán adatok esetében a *Homo sapiens* GRCh38, egérnél pedig a *Mus musculus* GRCm39 volt (Ensembl release 110, (Martin, F. J. *et al.*, 2023). Mindkét esetben a „toplevel” referenciát alkalmaztam, amely tartalmaz minden, a kromoszómákon nem elhelyezett, de a nukleáris genomhoz tartozó szekvenciát is.

Az első szűrési lépés volt azoknak a readeknek a kiválasztása, amelyek mitokondriális szekvenciára illeszkednek. Annak ellenére, hogy tisztított mitokondriumokból történt a szekvenálás, a readeknek jelentős része illeszkedett nukleáris genomi szekvenciákra. A mitokondriális szekvenciákra legalább részben illeszkedett leolvasásokat a samtools v1.18 (Li, H. *et al.*, 2009; Danecek *et al.*, 2021) szoftver használatával választottam ki.

A „tisztá”, azaz teljes egészében mitokondriálisnak adódott leolvasások szűréséhez használt módszert a „read ID” és referencia szekvencia párosokra alapoztam. A „read ID” egy leolvasás azonosító, amely egy egyedi, a bam formátumú fájlokban a leolvasás minden illeszkedésénél megtalálható karakter szekvencia (string). Amennyiben egy leolvasás mellett megtalálható volt az SA (supplementary alignment, kimérikus leolvasás jelenlétét jelzi) címke, az ellenőrzés következő lépéseként a read ID-hoz tartozó referencia szekvenciák azonosítóját vizsgáltam meg.

Amennyiben ezek közül az illeszkedések közül nem mind a mitokondriumra esett, az adott leolvasást teljes egészében kizártam a további elemzésből.

Az általam írt és alkalmazott szűrő scriptet ([M5. Melléklet](#)) a következő módon futtattam:

```
./select_pure_mt.py -i HP202.mt_aligned.bam -o ../5_cleaned/HP202.cleaned_mt.txt -b  
../5_cleaned/HP202.cleaned_mt.bam > ../5_cleaned/HP202.cleaning.log
```

```
./select_pure_mt.py -i BH10.mt_aligned.bam -o ../5_cleaned/BH10.cleaned_mt.txt -b  
../5_cleaned/BH10.cleaned_mt.bam > ../5_cleaned/BH10.cleaning.log
```

Az output minden esetben két fájl: egy bam fájl, amely a tisztán mitokondriális leolvasásokat tartalmazza, és egy egyszerű szöveges (txt) fájl, amely a read ID-ket sorolja fel. A script megbízhatóságát az „éles” futtatás előtt rövidített minta fájlkon teszteltem, amelybe véletlenszerűen választottam be leolvasásokat, ügyelve arra, hogy azok között legyen olyan, amely teljes egészében mitokondriumra illeszkedett, és olyan is, amely részlegesen, vagy egyáltalán nem. A teszteléshez a randomizált leolvasás szettet a samtools `--subsample 0.05` opcióval generáltam, ezen felül vizuális alapon választott pozitív és negatív kontroll leolvasásokat is adtam a fájlhoz. A szűrési eredményeket az IGV-vel (Thorvaldsdottir *et al.*, 2013) vizuálisan is ellenőriztem.

A szűrést követően a kapott read ID lista alapján az eredeti fastq fájlból a seqtk toolkit v1.4. (<https://github.com/lh3/seqtk>) subseq moduljával választottam ki a további elemzéshez felhasznált readeket.

4.3.4. Mitokondriális haplotípusok közötti megbízható különbségek azonosítása

A teljes egészükben mitokondriálisnak bizonyult leolvasásokat ezt követően ismét a pbmm2-vel (v1.13.0) (<https://github.com/PacificBiosciences/pbmm2>) illesztettem referenciára. Ekkor azonban már kizárólag mitokondriális szekvenciákat használtam referenciaként. A kevesebb adat és rövidebb referencia használata egyrészt meggyorsította ezt a lépést, másrészt ebben a lépésben referenciaként a sejtvonalak egyik haplotípusával megegyező mtDNS szekvenciát használtam. Ez a BH10 sejtvonala esetében a wt mtDNS volt, a HP202B esetében pedig az NZB mtDNS.

Több genetikai variáció hívó (variant caller) szoftver tesztelése után végül a GATK ToolKit (v.4.4.0.0) (Auwera és O’Connor, 2020) HaplotypeCallere mellett döntöttem. A ploiditást diploidként adtam meg, abból a megfontolásból, hogy körülbelül azonos mennyiségű mtDNS-re számítottunk a két haplotípusból: így várhatóan a diploidként hívott variációk közül a heterozigóták voltak a számunkra informatívak. A homozigóta variációk ennek megfelelően nem voltak informatívak, így a továbbiakban nem foglalkoztam velük. A rekombináns leolvasások azonosítása előtt pedig ezeket ki is szűrtem az ehhez használt vcf fájlból. A HaplotypeCallert alapértelmezett beállításokkal futtattam.

A BH10 sejtvonalon ezen kívül a pbsv v2.9.0 (<https://github.com/PacificBiosciences/pbsv>) szoftverrel strukturális variáció hívást (structural variant call) is végeztem, a Δ mtDNS-ben található deléciónak jelenlétének és pontos pozíciójának ellenőrzésére. A pbsv-t először „discover” módban futtattam, amely azonosítja azokat a leolvasásokat, amelyek strukturális variációk jelenlétére utalnak. A program ezt egy „svsig.gz” fájlba menti, majd a „call” mód ennek felhasználásával azonosítja a variációkat.

Feltételezéseim helytállóságát többször ellenőriztem vizuálisan is, az IGV használatával.

4.3.5. Rekombináns szekvenciák azonosítása – az alkalmazott script és működése

A rekombinációt azonosító script inputja az előkészített vcf fájl, és az illesztett leolvasásokat tartalmazó bam fájl. A vcf fájl sztenderd vcf formátumú, az egyetlen kritérium, hogy megbízható változatokat tartalmazzon. A bam fájl a megfelelő referenciára illesztett mitokondriális eredetű leolvasásokat tartalmazza.

A script a CIGAR tag alapján kalkulálja ki a leolvasási szekvencia egyes nukleotidjainak pozícióját a referencia szekvencián. Ezt hasonlítja össze az adott referencia pozíción található variációkkal, amelyeket a vcf fájl tartalmaz. A polimorf pozíciókra eső nukleotidok egy-egy referencia („A”, a BH10-wt, vagy az NZB referenciával megegyező) vagy nem-referencia („B”, BH10-Δ, vagy Balb/c típussal megegyező) jelölést kapnak. Ezáltal egy A és B jelölésekből álló pszeudo-szekvencia jön létre.

A fentebbiek alapján létrejött szekvencia kiértékeléséhez ez egyszerű pontozási rendszert hoztam létre:

- Amennyiben egyetlen SNP tér el az azt körülvevő szekvencia típusától, -1 pontot jelent
- Egynél több, de túl kevés SNP-t tartalmazó eltérés 0 pontot ér
- Minden szakasz, amely egy megadott minimumnál hosszabb, +1 pontot ér

A script futtatásakor a minimálisan pozitívan értékelt szakasz hossza beállítható a **--minstretch** opcióval, a **--minscore** opcióval pedig a minimum elérendő pontszám állítható be. Amennyiben egy read megkapja a minimum pontszámot, az pozitív találatként jelenik meg a kimeneti fájlokban.

A fő kimeneti fájl egy tab szeparált táblázat (tsv), amely tartalmazza a leolvasás azonosítóját, az A és B jelölésekből képzett „pszeudo CIGAR” stringet, és az eredményt, True/False formában. A negatív találatok kiírása opcionális, az **--out_negative** opció segítségével kérhető. Az említett „pszeudo CIGAR” a bam fájlok általánosan alkalmazott CIGAR stringjének elvén működik, a detektált A és B típusú pozíciók számát összesíti. Például egy „AAABBBAAAA” szekvencia „3A3B4A”-ként kerül kiírásra. A kimeneti fájl tartalmazza ezen felül a deléciós típust is, ez azonban csak a BH10 sejtvonalnál hordoz tényleges információt.

A fentiekén kívül megadható a **--plotdata** opció is, amelyet a plottolás megkönnyítése érdekében tettem hozzá a scripthez. Az opció nem automatizálja a plottolást, egyszerűen olyan formába szervezi az eredményt, amely azt megkönnyíti – szemben az olvashatóságot szem előtt tartó, korábban leírt fő kimeneti fájllal. A vizualizációt egy különálló R scripttel oldottam meg.

A detektáló scriptet kísérleteim folyamán többféle beállítással is lefuttattam. A publikált adatokhoz végül a **--minscore 3** és **--minstretch 2** beállításokat választottam, amely a BH10 sejtvonala esetében a legtöbb, valóban megbízható találatot adta.

Valamennyi használt script elérhető a <https://zenodo.org/records/10473729> linken a cikk publikációjának idejében használt formában, valamint ezen dolgozat [M5. mellékletében](#).

5. EREDMÉNYEK ÉS DISZKUSSZIÓ

5.1. Házi- és üregi nyúl hibridizációja, valamint szelekció a nyúl háziasítása során

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

5.1.1. Szekvenálási eredmények és minőségi szűrés

Összesen 98 nyúl minta teljes genom szekvenálási adataival dolgoztam, amelyből 56 más tanulmányokból származó, publikusan elérhető, 42 pedig újonnan, saját munkámhoz szekvenált minta volt. Az újonnan szekvenált minták között volt 2 mintavételi helyről 14 magyarországi üregi nyúl minta, valamint minták a Rex, Tübingiai, Kaliforniai, Hycote, Zika és Új-Zélandi fehér fajtákból. Az NCBI SRA-ról letöltött minták között Új-Zélandi fehér, Japán fehér, Watanabe, holland törpe, és angora fajtákból származó egyedi minták mellett poolozott minták is voltak, francia és ibériai üregi nyúl populációkból, valamint belgian hare, holland (Dutch), flamand óriás (Flemish giant), francia lógófülű (French lop), és Champagne d'Argent fajtákból. Az általunk szekvenált genomok elérhetőek az NCBI SRA adatbázisában, BioProject PRJNA1118464 azonosító alatt. A felhasznált minták teljes táblázata az [M2. mellékletben](#) található.

Saját mintáinkhoz képest a publikus minták lefedettsége jóval alacsonyabbnak bizonyult. Ez nagyobb hátrányt jelentett volna, ha az összes felhasznált szekvenálás lefedettsége hasonlóan alacsony. A nagy lefedettségű szekvenálások, amelyekből jó minőségű, megbízható variációkat tudunk hívni az együttes genotipizálás során, segített a variációk konszolidálásában, csökkentve ezt a hatást.

A minőségi ellenőrzés, referenciára illesztés és variációk hívása után kb. 27,5 millió variációt kaptam. Az ez után következő valamennyi szűrési feltételt 2 780 336 variáció teljesítette. Az egyes szűrési lépések után megtartott variációkat az **1. táblázat** foglalja össze.

1. táblázat - Variációk az egyes szűrési lépéseket követően. Az oszlopok a szűrési lépéseket, a sorok a teljes megtartott mintaszámot (Minta), az ebben megtalálható összes variációt (Variáció), illetve ezek típusait mutatják.

Szűrési lépés	Minőségi szűrés	Genotipizáltság és allél szám	LD pruning
Minta	100	97	97
Variáció	67 084 502	27 524 641	2 780 336
SNP	50 253 306	21 863 765	1 825 434
Indel	18 882 395	5 660 876	954 902
Egyéb (SNP és Indel)	1 352 355	0	0
Multiallélikus	8 344 576	0	0
Multiallélikus SNP	2 506 463	0	0

A 2009-ben megjelent OryCun2.0 referenciát (NCBI RefSeq assembly GCF_000003625.3, GenBank accession GenBank accession GCA_000003625.1) csupán 2024-ben váltotta fel az új, mOryCun1.1 referenciagenom (NCBI RefSeq assembly GCF_964237555.1, GenBank accession GCA_964237555.1). A korábbi, az OryCun2.0 referencia összeszereléséhez felhasznált

szekvenálás beltenyészett, és ebből adódóan magas homozigotitású, a Thorbecke új-zélandi fehér vonalból származó nőstény egyedből készült (Carneiro *et al.*, 2014b). A magas homozigotitás, és az összes szekvencia diploid jelenléte előnyös volt az ekkor elérhető összeszerelési technikák szempontjából, viszont a nőstény egyed felhasználása egyben azt jelentette, hogy Y kromoszóma szekvencia nem állt rendelkezésre. Az OryCun2.0 3317 scaffoldból áll, amely 84023 contigból tevődik össze.

Összehasonlításképp, a korábbi referenciával (OryCun2.0, 2009) szemben az mOryCun1.1 (2024) nagy pontosságú hosszú reades szekvenálási technikával (PacBio HiFi) készült, a scaffoldoláshoz pedig Hi-C szekvenálási adatok kerültek felhasználásra. Ez az összeszerelt genom 224 scaffoldból és 2497 contigból áll, továbbá tartalmazza az X és Y kromoszómákat is.

Mivel a GATK által ajánlott szűrési munkafolyamat eredendően humán mintákra optimalizált, illetve a rendelkezésre álló jó minőségű, és kiterjedt humán adatbázisok igénybevételére alapul, a VQSR-t, amely egy gépi tanulás alapú minőségi szűrési módszer, nekem nem volt célszerű lefuttatnom. Nyúlhoz nem állnak rendelkezésre az dbSNP-hez (Phan *et al.*, 2025; <https://www.ncbi.nlm.nih.gov/snp/>) hasonló adatbázisok, amelyek ehhez ajánlottak. Ezen felül voltak mintáim között fajták, illetve populációk, amelyekből 1-1 egyed állt rendelkezésre, így az azokban található egyedi variációkat sem volt célszerű eltávolítani. A szűrést így egyszerűen számszerű minőségi jellemzők alapján végeztem, a szűrési feltételeket pedig úgy választottam meg, hogy a lehető legmegbízhatóbb variációk maradjanak bent, a lehető legnagyobb változatosság megtartása mellett.

A nagy sűrűségben hívott, jó minőségű variációk javítják a további vizsgálatok során a detektálás érzékenységét (James *et al.*, 2022). A legtöbb minta esetében meglévő magas lefedettségnek köszönhetően kiszűrhettem minden olyan hívott változatot, ami akár a legkisebb mértékben is kétséges minőségűnek tűnt. A GATK sztenderd jó gyakorlatát követő genotipizálás és szűrés után is több millió variáció áll rendelkezésemre, további, az egyes tervezett analízisekhez szükséges szűréseket lehetővé téve. Bár a GATK ToolKit eszközeinek alapvető beállításai általában humán adatokra optimalizáltak, az emlős genomok alapvető hasonlóságainak köszönhetően üregi nyúl szekvenciák esetében is jól működnek ezek a beállítások.

Az ekkor rendelkezésre álló OryCun2.0 referenciagenom jellemzője volt a nagy mennyiségű, kromoszómákon nem elhelyezett, vagy azokon nem lokalizált szekvencia jelenléte. A kromoszómákon nem lokalizált szekvenciák kromoszómákhoz rendelték, de azokon belüli elhelyezkedésük nem ismert. Mivel számos bioinformatikai program nem tudja kezelni az ilyen szakaszokat, végül azokat kihagytam az elemzésből. Ezzel egyes gének, vagy más, szabályozási szempontból fontos szakaszok is kimaradhattak, amelyek kezelése az mOryCun1.1-gyel már nem jelentene problémát.

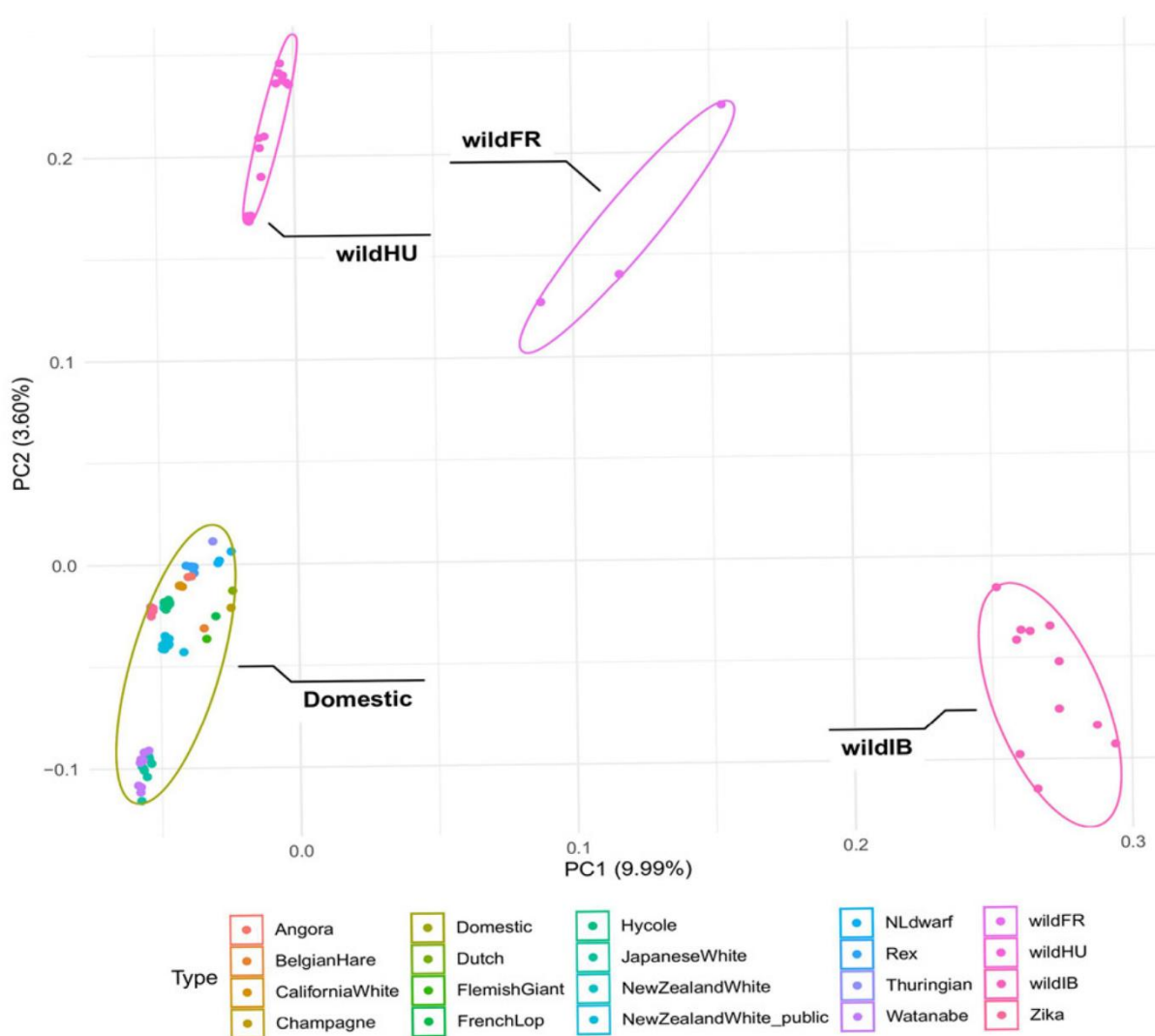
A rendelkezésre álló adatmennyiség miatt megtehettem, hogy kizárólag olyan hívott variációkat tartok meg a végső szettben, amelyek minden mintában megbízhatóan genotipizált pozíciókon találhatóak.

5.1.2. Populáció-struktúra

A főkomponens-elemzés eredményei (**2. ábra**) alapján az üregi- és házinyúl populációk tisztán elkülönülnek. Ezen belül is jelentősnek adódott a különbség a különböző földrajzi területekről származó üregi nyúl populációk között. Kisebb mértékben, de tisztán elkülönülnek az egyes házinyúl fajták is (**3. ábra**). A házinyúl fajták közötti legnagyobb különbség a két japán fajta (japán fehér, és az ebből kitenyészett Watanabe) elkülönülése a többi, alapvetően európai és észak-amerikai eredetű fajtától. Ez a két fajta a filogenetikai fán is messzebb esik a többitől (**4. ábra**).

A maximum-likelihood módszerrel kalkulált filogenetikai fa érdekessége, hogy míg a PCA-n az ibériai és francia üregi nyulak teljesen külön csoportot alkotnak, a fán a két populáció azonos ágon helyezkedik el, mutatva a két populáció közös eredetét. Ezen belül a francia üregi nyúl populáció relatív közelsége a háziyúl fajtához egybevág Carneiro és munkatársai (Carneiro *et al.*, 2011) korábbi következtetéseivel. Eszerint az üregi nyúl az Ibériai-félszigetről terjedt Franciaország területére, ahol a háziásítás és fajtanemesítés megkezdődött.

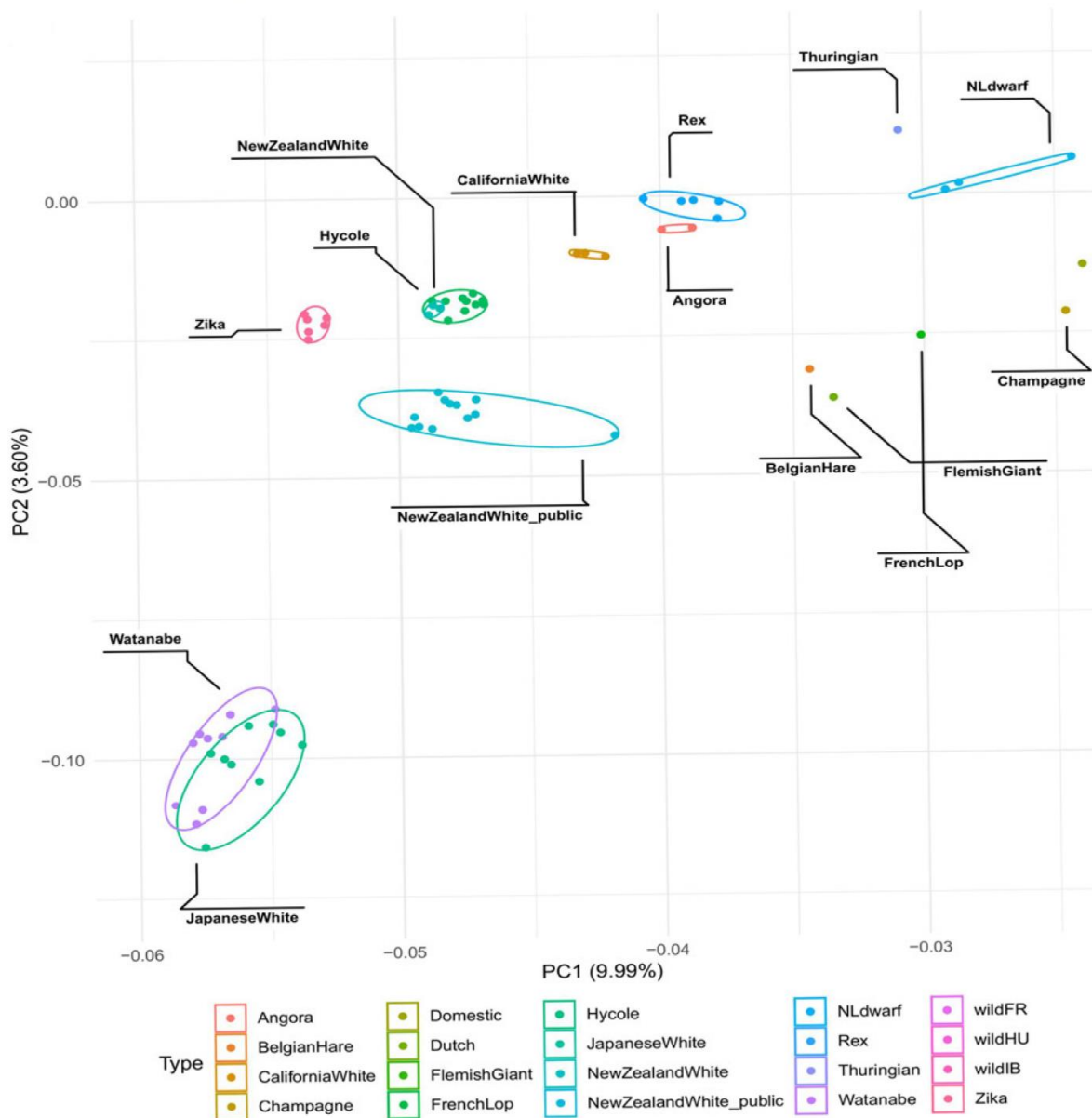
A francia és ibériai üregi nyúl populációkkal azonos eredettel, de elkülönülő ágon találjuk a magyarországi üregi nyulakat. A két magyar mintavételi helyszín ezen az ágon nem elkülöníthető. Mindezek alapján a magyarországi üregi nyúl állomány vélhetően a háziásítás előtt vált el az ibériai eredetű populációktól. Az, hogy a faj emberi közvetítéssel, vagy önállóan terjedt el a Kárpát-medencében, nem ismert, de a genetikai távolság és a filogenetikai fán való elhelyezkedés alapján valószínűsíthetjük, hogy az állatok nem elvadult háziyulak leszármazottai.



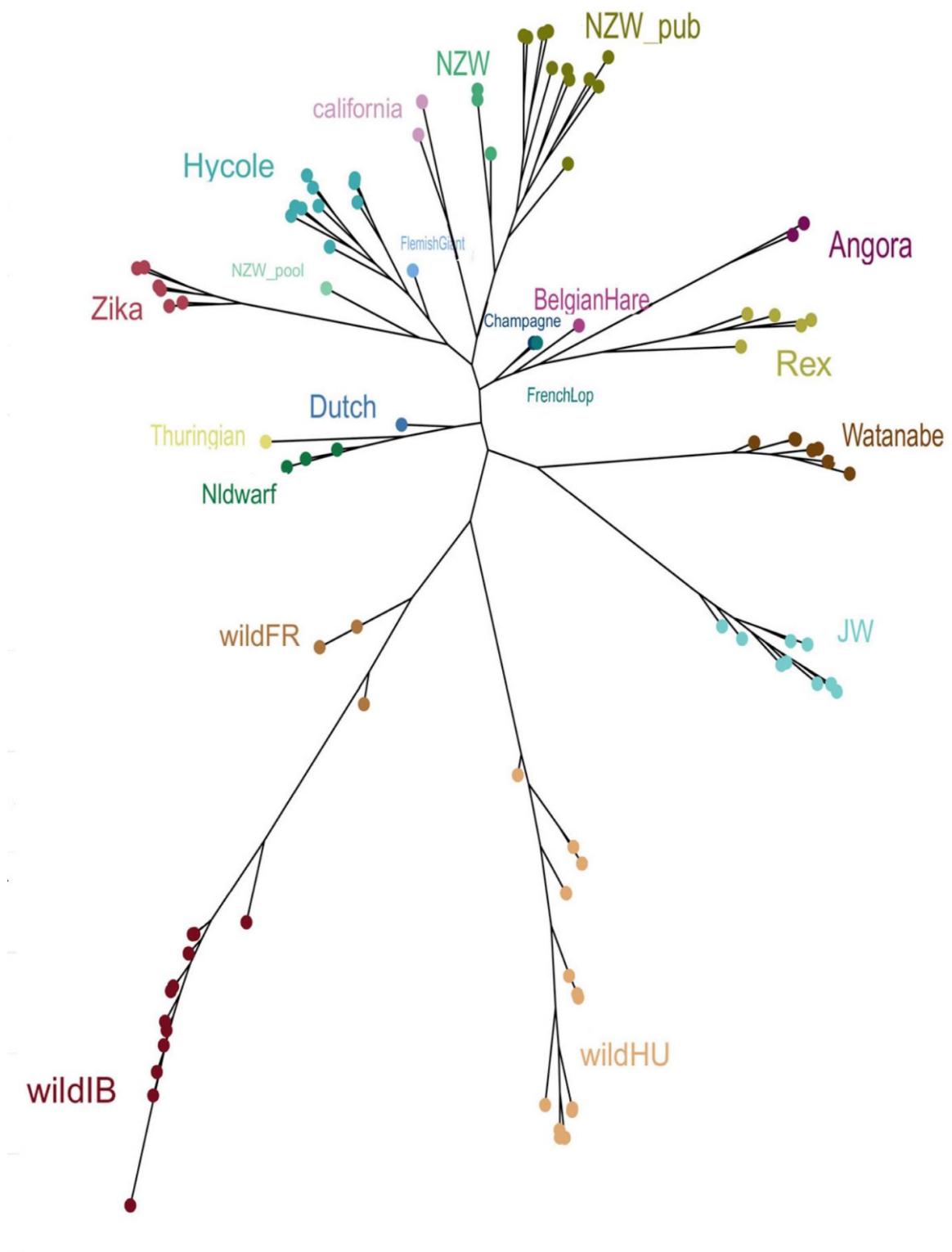
2. ábra - Az összes minta főkomponens analízise (PCA). Az egyes vad üreginyúl populációk (wildHU, wildFR, wildIB – magyarországi, franciaországi, és ibériai-félszigeti) jól láthatóan elkülönülnek a háziaktól (Domestic). (Fekete *et al.*, 2025b)

A háziyúl fajták általában a fajta elsődleges felhasználása alapján kerülnek egy csoportba a fán: az Új-Zélandi fehér, Kaliforniai, Hycole és Zika nyulakat főként hústermelési céllal tenyésztik, míg a Rex és Angora nyulakat, bár népszerűek házikedvencként is, eredetileg szőrme- és

gyapjútermelési céllal tenyésztették ki. Ehhez a két célhoz a szőr eltérő minőségi tulajdonságai szükségesek, de mindkét esetben a szőrképzés komplex genetikája és szabályozása játssza a kulcsszerepet. A szőrképzés komplexitásából adódóan, bár ismertek hozzá kapcsolt gének, a fajtákra jellemző tulajdonságok kialakulásának pontos genetikai háttere és mechanizmusa nem ismert egyik esetben sem (Dai, Y. *et al.*, 2022; Fatima *et al.*, 2023). Egymáshoz közel csoportosulnak a Tübingiai (Thuringer), holland törpe (Netherlands dwarf), és holland (Dutch) fajták is, melyek mind kistestű fajták, és elsősorban házikedvencként tartják őket. Annak ellenére, hogy a törpe (dwarf) fenotípust a holland törpenyúl esetében egy ismert delécio okozza, a fajta nem-törpe egyedei is kis termetűek, amely más, testmérettel kapcsolt gének befolyására utal (Carneiro *et al.*, 2017).



3. ábra - A főkomponens analízis (PCA) házinyulakat tartalmazó csoportja. A Watanabe és japán fehér (JapaneseWhite) fajták külön csoportot alkotnak a többi, főként európai és észak-amerikai házinyúl fajták alkotta csoporttól. Érdekes módon az egyaránt az új-zélandi fehér fajtába tartozó, általunk szekvenált (NewZealandWhite) és a publikus adatbázisból letöltött (NewZealandWhite_public) minták is jól elkülöníthető csoportot alkotnak, az egyes fajtákon belül jelenlévő populáció struktúráját mutatva. (Fekete *et al.*, 2025b)



4. ábra - Filogenetikai fa maximum-likelihood módszerrel. A házinyúl fajták jobbra 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 fán található egy, az azonos fajtaból származó mintákhoz képest inkonzisztensen elhelyezkedő minta is, ez az NCBI SRA-ról letöltött, publikusan elérhető, poolozott, Új-Zélandi fehér fajtájú minta (SRA accession SRS609672), amely eredményeim alapján közelebb esik a Zika fajtájú mintákhoz. Ennek több oka lehet, többek között a fajtákon belüli, és egyes tenyészetek közötti erős strukturáltság, esetlegesen a Zika és Új-Zélandi fehér fajtájú állomány hibridizációja. Mivel nem áll rendelkezésemre részletes információ a minta eredetéről, és a poolban található egyedekről, a konkrét okot nem lehetséges egyértelműen meghatározni.

5.1.3. Hibridizáció vad- és házinyulak között

Az ADMIXTURE program keresztvalidálási (cross-validation, cv) eredményei alapján a felmenő populációk (K) legvalószínűbb száma K=5. Érdekes módon a kiugró K=6 után K=7-nél ismét alacsonyabb a keresztvalidálási hiba értéke. Ennek megfelelően K=6-nál a feltételezett eredeti populációk nagyban eltérnek a K=5 és K=7 értéknél számítottaknál. K=6 értéknél az algoritmus eggyel több üregi nyúl populációt különített el, míg a másik kettőnél inkább a házinyúl fajták váltak el. A számított hiba alapján a K=5 és K=7 értékek közelebb állnak a tényleges populáció struktúrához.

A hibridizáció mértékét K=5, K=6, és K=7 értékeknél is megvizsgáltam. Az ADMIXTURE által elválasztott populációk az alábbiak szerint alakultak: K=5 esetében a magyar üregi nyulak alkotnak egy csoportot, a francia és ibériai üregi nyulak együtt szintén egy csoportot alkotnak, míg a két japán házinyúl fajta elkülönül, és a többi házinyúl fajta alkotja az ötödik csoportot. K=7 esetében a házinyúl fajtákat több eredeti populációra bontja a program. Mindhárom vizsgált K érték potenciális hibridizációt mutat házinyúl és magyar üregi nyúl, valamint házinyúl és francia üregi nyúl populációk között (5. *ábra*).

A nyúl házasítása sok más haszonállathoz képest a relatív közeli múltban történt. Bár a genetikai különbségek egyértelműek, és jól elkülöníthető populációkat eredményeznek, és a házinyúl a domesztikáció és fajtaképzés során átment több erős palacknyak hatáson (bottleneck) is, az üregi nyúl és házinyúl populációk között nincs olyan jellegű szeparáció, amely gátolná azok keveredését.

Az Ibériai-félszigeten előforduló két alfaj, az *O.c.cuniculus* és az *O.c.algirus* között megfigyelhető korlátozott genetikai áramlás, illetve az X kromoszómán szignifikánsan kisebb arányban előforduló rekombináció - és ebből adódóan kevesebb genetikai változat introgressziója – figyelhető meg (Carneiro *et al.*, 2014a). Ennek jelenléte a házinyúl és üregi nyúl között, amelyek közül mindkettő az *O.c.cuniculus* alfajhoz tartozik, nem ismert. Nem kizárt, hogy a vadon élő populációkból egyes variációk, azok fitnessre gyakorolt hatása miatt kiszelektálódnának, ugyanakkor több fajban megfigyelték a házasított változathoz a vad típus irányába történő génintrogressziót (Besnier *et al.*, 2022; Goedbloed *et al.*, 2013; Heikkinen *et al.*, 2020; Mattucci *et al.*, 2019; Oliveira *et al.*, 2008; Pilot *et al.*, 2018; Verardi *et al.*, 2006). Egyes esetekben ezek a változatok pozitív szelekció alatt is állhatnak (Pilot *et al.*, 2021). Házinyulak kiszabadulása, és hibridizációja vad üregi nyúl állományokkal tehát befolyással lehet azok minőségére. A házinyúl és az üregi nyúl közötti genetikai áramlás befolyásának mértékének és hatásának felderítése jelentős mennyiségű további kutatást igényelne, amely meghaladja dolgozatom kereteit.



5. ábra - ADMIXTURE eredmények. $K=5$ és $K=7$ (ahol K feltételezett felmenő populációk száma) esetén kaptuk a legjobb keresztvalidálási értékeket. A két csoportosítás hasonló eredményt adott, azzal a különbséggel, hogy az Új-Zélandi fehér (NZW és NZW_public) és Zika minták elkülönülnek a többi házi fajtától. $K=6$ esetében az ibériai-félszigeti és francia minták mutattak erősebb strukturáltságot. Minden esetben hasonló mértékű hibridizáció feltételezhető a házinyulak és magyar, illetve francia üregi nyulak között. (Fekete et al., 2025b)

ADMIXTURE eredményeink alapján a franciaországi üregi nyulak jelentős, míg a magyarországiak kisebb mértékű hibridizációt mutatnak házinyulakkal. A két üregi nyúl populáció közül a magyar mintákat vizsgáltam tovább, amelyek mind nagy lefedettségű, egyedi szekvenálásokból származnak, szemben a poolozott, és relatív alacsony lefedettségű francia mintákkal.

A magyar üregi nyúl mintákat a mintavételi hely szerint (Tiszaszentimre környékén vadon élő, illetve a Budapesti Állat- és Növénykertben tartott üregi nyulak) két szubpopulációra bontva kiderült, hogy ezek közül csupán az egyik, az állatkerti populáció mutatja házinyullal történt hibridizáció jeleit. Treemix szoftverrel végzett elemzéseim is alátámasztják ezt az eredményt: 0 és 5 közötti hibridizációs eseményt feltételezve konzisztensen adódik egy az állatkerti üregi nyúl és a Thüringiai fajta között (**6. ábra**). Ugyanakkor összesen 1 Thüringiai fajtájú minta áll rendelkezésemre. Az ADMIXTURE eredmények alapján a Thüringiai alapvetően a többi házi fajtához erősen hasonló genetikai háttérrel rendelkezik, így az, hogy a Treemix szoftver ezzel a fajtával mutatott ki hibridizációt, nem konkluzív. D és f4 statisztikákkal megerősítve (**M6. Melléklet**) elmondható, hogy az állatkerti üregi nyúl populációban találhatóak házinyulakkal hibridizált egyedek, de a bekeresztett fajta, illetve fajták nem állapíthatóak meg egyértelműen.

Az egyes állatkerti egyedek visszakövetésére nem volt módomban, így kérdéses marad az is, hogy a hibridizáció hol, mikor és hogyan következhetett be.

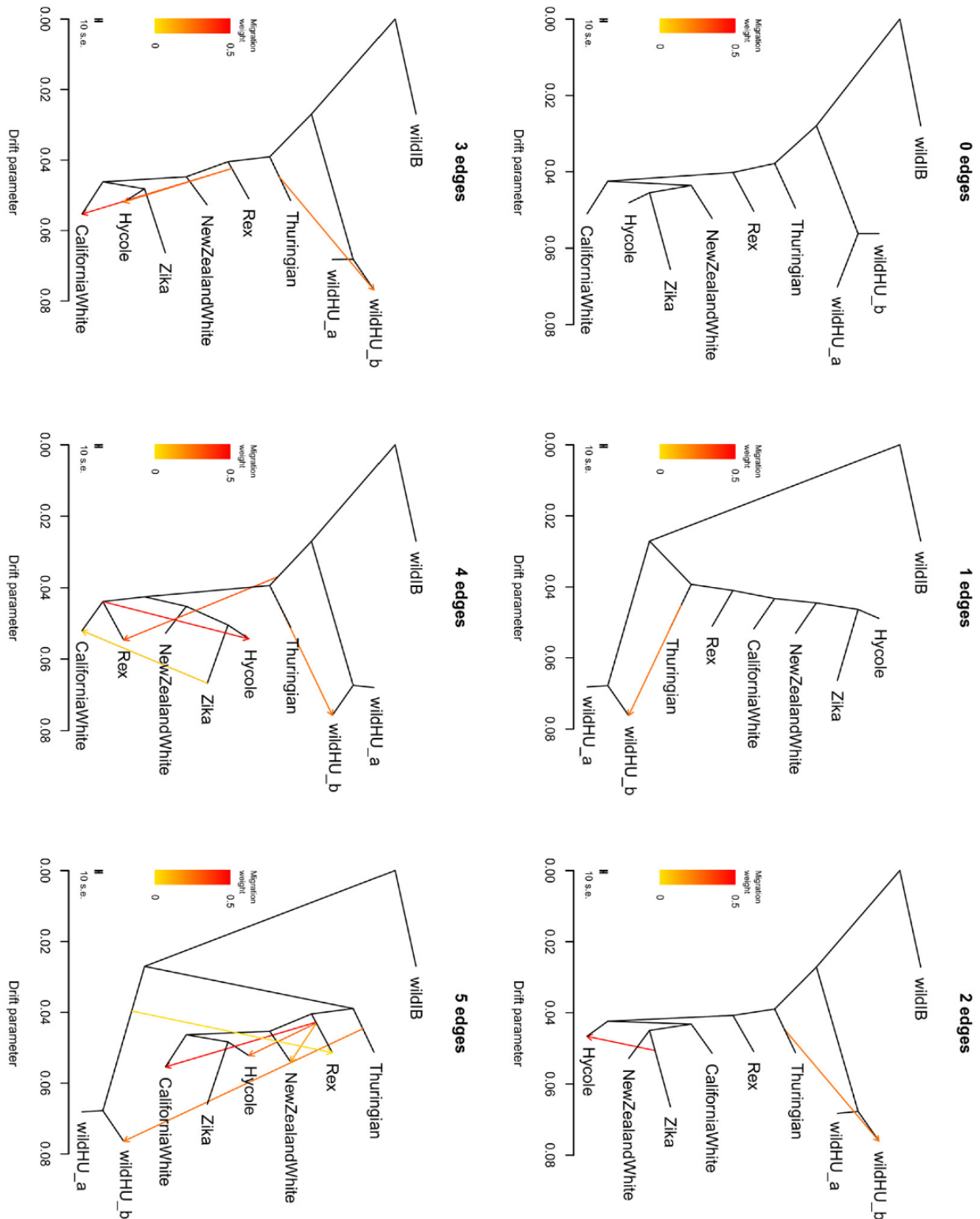
5.1.4. A beltenyésztettség mértéke

Az ibériai és francia üregi nyúl minták esetében nem volt lehetőségem megvizsgálni, és a többi populációval összevetni a homozigotizáció mértékét, valamint egyes publikus házinyúl mintákat is ki kellett hagynom ebből az elemzésből. Ennek oka, hogy a poolozott minták genotipizálásánál a heterozigotizáció túlbecsült lesz, mivel nem elkülöníthetőek a poolban található egyes egyedekhez tartozó leolvasások, és a hívó algoritmus több, eltérő allélre homozigóta egyedeket könnyen egy heterozigóta genotípusként értelmezhet (Lynch *et al.*, 2014). Az elemzésből kizárt minták: a PRJNA242290 és PRJNA354575 SRA BioProject ID-hoz tartozó minták (**M2. Melléklet**).

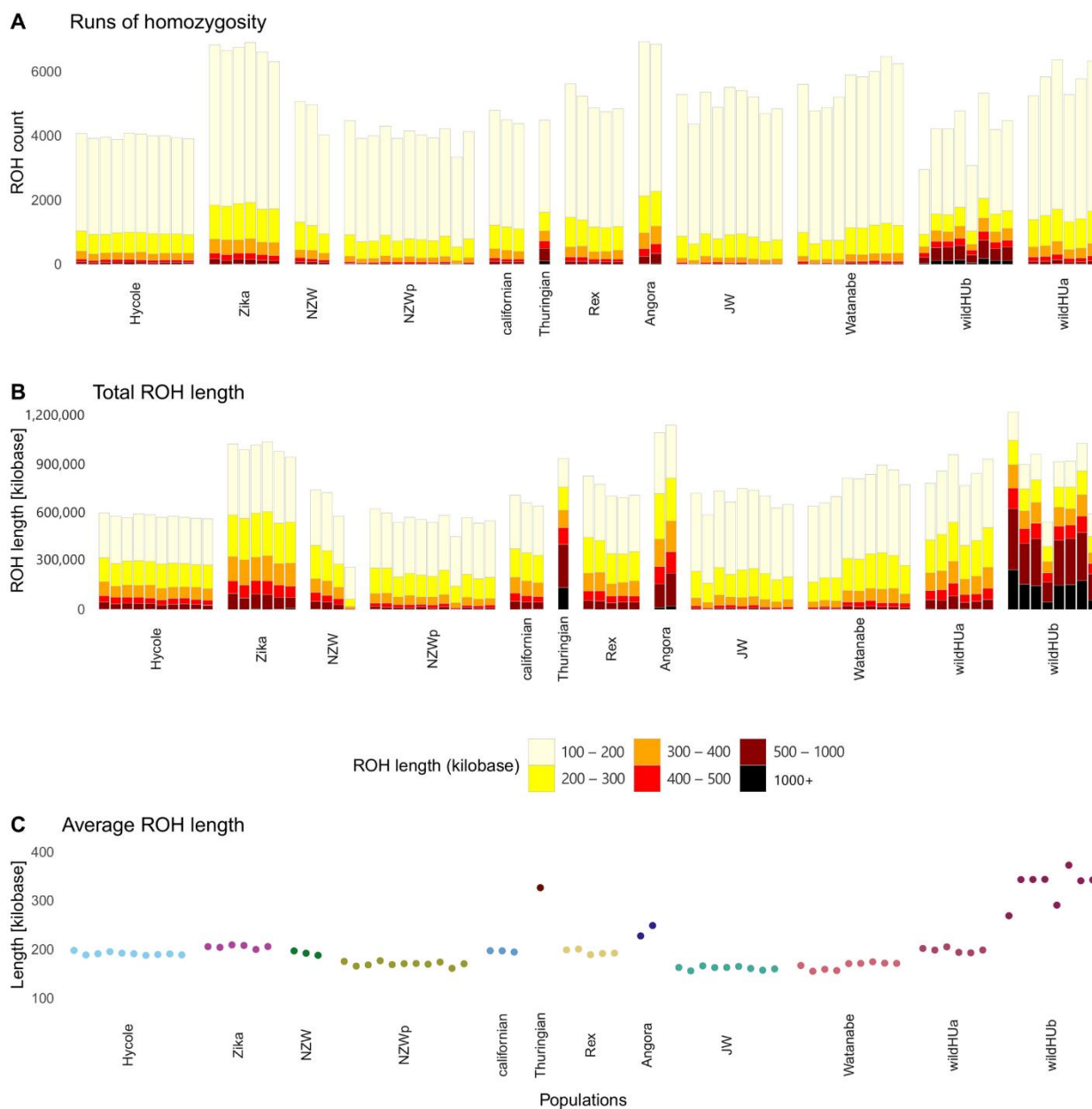
ROH elemzéseknél meghatározó kérdés, hogy milyen hosszúságú szakaszt számítunk ROH-ként, illetve, hogy megengedünk-e egy-egy heterozigótaként hívott változatot egy ROH-ban. Kisebbségi felbontású adatok, azaz ritkábban elhelyezkedő variációkat tartalmazó minták esetén is eltérő értékeket kaphatunk a nagy felbontással vizsgált egyedektől. Vizsgálatomhoz az összes egyedileg szekvenált mintát fel tudtam használni, mivel minden esetben teljes mértékben ugyanazokon a genetikai koordinátákon elhelyezkedő variációkat használhattam. A nagy mennyiségű variáció és nagy felbontás miatt aránylag rövid (100 kb) hosszúságú homozigóta szakaszokat már figyelembe vettem az elemzésnél. A különböző hosszúságú szakaszok számát és összesített hosszát vizsgált egyedenként a **7. ábra** mutatja.

A homozigóta szakaszok (ROH, run of homozygosity) hossza alapján a két magyarországi üregi nyúl populáció eltérő mértékű beltenyésztettséget mutat. A Budapesti Állat- és Növénykertben tartott populáció a közeli múltban történt beltenyésztettség jeleit mutatja, kevesebb, de hosszabb ROH régióval. Ebben a populációban volt a homozigóta szakaszok teljes hossza a legnagyobb, amelyet a Zika és Angora nyúlfajták egyedi követnek. Kiemelkedően hosszú ROH régiókat találtunk a Thüringiai fajtában is. Érdekes módon ugyanez, a beltenyésztés jeleit mutató állatkerti populáció adódott hibridizáltnak házinyulakkal. Az állatkerti populáció kis létszáma miatt abban könnyen kialakulhat beltenyésztés, még akkor is, ha esetlegesen az állományhoz adnak egy-egy új példányt. A házinyullal való hibridizáció idejéről sajnos nincs tudomásunk, ennek specifikus

vizsgálatára nem volt alkalmam, így nem tudjuk, hogy az az állatkerten belül került a populációba, vagy a befogott állatok már hibridek voltak.



6. ábra - Treemix szoftverrel kapott eredmények. A populációk között bekövetkezett feltételezett hibridizációs események számától függetlenül az egyik magyar populáció (wildHU_b) hibridizációt mutat házinyulakkal, amely egybevág az ADMIXTURE elemzés eredményeivel. A hibridizációs folyamatok irányát nyilak mutatják, míg a nyilak színe annak mértékét jelzi, a sárgától a piros felé erősödő skálán. (Fekete et al., 2025b)



7. ábra – Homozigóta szakaszok (ROH) a vizsgált házinyúl és üregi nyúl populációkban. **A)** ROH szakaszok száma, az egyes szakaszok hossza szerint színezve. **B)** ROH szakaszok teljes, kumulált hosszúsága, az azokban található egyes ROH hosszúságoknak megfelelően arányosan színezve. Míg a ROH-ban található genomi szakaszok teljes hossza nem mutat jelentős különbséget, a wildHU_b (Budapesti Állat- és Növénykert) magyarországi üregi nyúl populációban ezek hosszabb szakaszokban helyezkednek el, ami a közelmúltban történt beltenyésztésre utal. **C)** Átlagos ROH hosszúság egyedenként. A wildHU_b populáció és a türingiai minta itt is kiemelkedő értékeket mutat. (Fekete et al., 2025b).

A Tiszaszentimre környékén vadon élő üregi nyulakban nem volt jelen az átlagosnál több, vagy rendkívül hosszú ROH, és házinyulakkal sem hibridizáltak az általam végzett tesztek alapján.

5.1.5. Szelekció által érintett régiók

A szelekciót kizárólag olyan egyedekben vizsgáltam, amelyek nem mutatták a házinyúl és üregi nyúl hibridizációjának jeleit. Ez alapján 4 magyar és az összes francia üregi nyúl mintát kizártam az elemzésből.

A szelekció fő jelei az adott genomi régióban a csökkent nukleotid diverzitás és az eltérő allélfrekvenciák által mutatott magasabb differenciációs szint. A tesztek kiválasztásánál a módszerek biológiai alkalmassága mellett figyelembe vettem azok alkalmazhatóságát a teljes genom szekvenálásból származó nagy mennyiségű adat feldolgozásához, skálázhatóságát és erőforrás-igényét is. A választott három módszer könnyen beállítható hasonló léptékre csúszóablakos megoldással. A felbontás és a számítógépes erőforrás-használat közötti kompromisszumként döntöttem a 10 kbp hosszúságú csúszóablakok mellett, 5 kbp hosszúságú lépésekkel, amellyel mindhárom módszer elfogadható sebességgel működött.

A szelekció vizsgálatára kiválasztott három teszt közül az első, a nukleotid diverzitás (π) a populáción belüli polimorfizmus mértékét méri. A szelekció alatt álló változatok frekvenciájának változásával – azok elterjedésével vagy eltűnésével – a nukleotid diverzitás mindenképp csökkenést mutat.

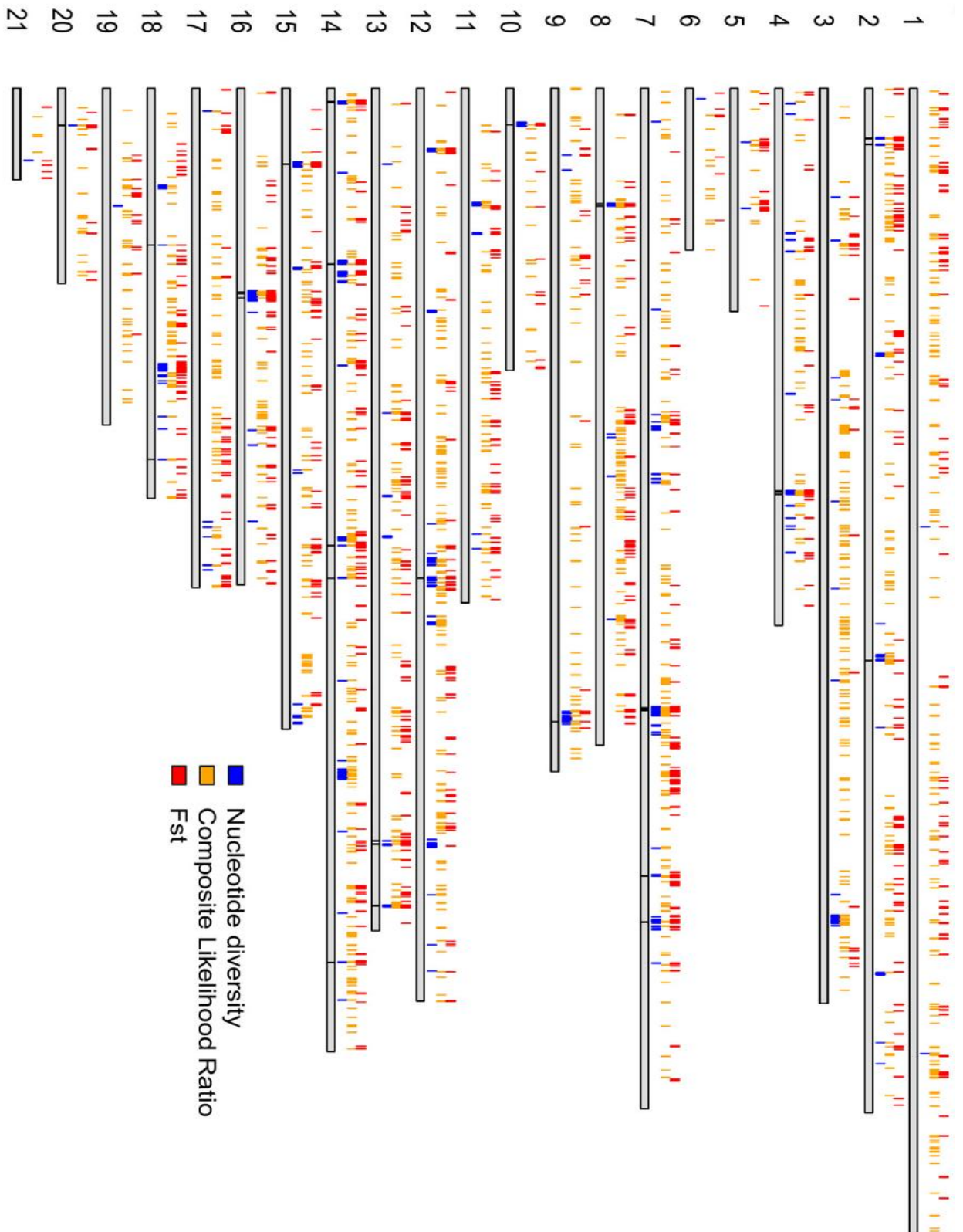
A fixációs index (Fst) az egyes populációk közötti különbséget számszerűsíti az allélfrekvenciák különbsége alapján. A 0-1 közötti érték minél közelebb esik az 1-hez, annál inkább differenciált az adott lókuszon a két populáció.

A harmadik, a SweepFinder2-ben (DeGiorgio *et al.*, 2016) implementált Composite Likelihood Ratio teszt (CLR) az allélfrekvencia-spektrum vizsgálatán alapul. Első lépése az empirikus háttérfrekvencia kalkulálása. Az alternatív hipotézis, hogy ezt a semleges spektrumot egy szelektív sweep megváltoztatta. A SweepFinder2 lehetőséget ad rekombinációs térkép figyelembevételére, nyúlban azonban ilyen nem áll rendelkezésre.

*Az Fst és a CLR statisztikáknál a magas, míg a π esetében az alacsony értékek utalhatnak szelekcióra. Ezt figyelembe véve szűrtem le az eredményeket a tesztek 99. percentilisbe eső értékekre. Ezeknek a régióknak az átfedéseiből alakítottam ki a végső régiósztetet. Az egymás közelében legfeljebb 20 kbp-re található régiókat összeolvasztottam a könnyebb kezelhetőség érdekében. Ilyen módon 46, a domesztikációban potenciálisan szelektált régiót határoztam meg, amely mindhárom módszerrel kielégítette a szűrési feltételeket. Az egyes módszerekkel szelektáltak adódó régiókat és elhelyezkedésüket a **8. ábra** mutatja, a szelektált régiókat és a bennük található géneket pedig a **2. és***

3. táblázat foglalja össze, míg az ezekben található, erősen differenciált változatokat az [M3. melléklet](#) részletezi.

A kromoszómákon összesen 1,985,753 feltételezett transzkripció faktor kötőhelyet azonosítottam a ChipSummitDB adatbázis adatai alapján. Ezek közül 1170 esett potenciálisan szelektált régiókba ([M4. Melléklet](#)), de ezek közül csak 93 olyanba, amely nem fed át egyetlen génnel sem. A Fisher-exact teszt alapján szignifikáns dúsulás mutatkozott 7 különböző típusú transzkripció faktor kötőhely esetében (**4. táblázat**).



8. ábra – Különböző módszerekkel szelektáltak adódó régiók elhelyezkedése az autoszómákon (szürke sávok). A szürke alapon feketével jelzett régiók a három módszer átfedését jelzik, míg az egyes kromoszómák feletti színes sávok az egyes módszerekkel egyesével szelektáltak feltételezhető régiókat jelölik. (Fekete et al., 2025b)

Feltételeztem, hogy a domesztikációval járó változásokban nem csak a génekben, hanem azok expressziójának szabályozásában részt vevő szakaszokban található mutációk is szerepet

játszanak, amelyre Sato és munkatársai (2020) transzkriptomikai vizsgálata is utal. A transzkripció faktor kötőhelyek azonban közvetlenül nyúlban nem kerültek eddig térképezésre, így a nyúl és humán genom hasonlóságait kihasználva LiftOver módszerrel azonosítottam olyan TFBS-eket, amelyek funkciója nagy valószínűséggel azonos. A TFBS pozíciókon azonosított kevés szelektált variáció oka lehet részben az azonosításra alkalmazott megoldás: a rövid genetikai szakaszok mutációi könnyen megakadályozzák azok felismerését. A módszerrel a ChipSummitDB-ben megtalálható humán TFBS-ek mennyiségének töredékét tudtam nyúlban azonosítani.

A génontológiai tesztek FDR-korrigált eredményei alapján nem volt szignifikáns dúsulás egyetlen vizsgált kategóriában sem. Ugyanakkor kevés gén került bele a szigorú feltételek miatt.

Más publikációkban, ahol kevesebb feltételnek kellett a jelölt régióknak eleget tenni – például két módszerrel kapott eredmények átfedése elegendő volt –, több GO kategóriában is adódott szignifikáns dúsulás (Ballan *et al.*, 2022). A korábban házinyúlön végzett szelekciós vizsgálatok azonban általában fajták közötti szelekcióra, nem pedig a domesztikáció során törtéetre fókuszálnak (Ballan *et al.*, 2022, 2023; Li, C. *et al.*, 2022; Liu, C. *et al.*, 2021). A kísérletembe bevont változatos fajtáknak köszönhetően a fajtaspecifikus tulajdonságok általánosan szelektáltként történő detektálása csekély.

A természetből származó, vadon élő üregi nyúl minták, szemben a többszörös palacknyak hatáson átment házinyúl populációkkal, különböző környezetből származó, ahhoz alkalmazkodott állatok. A különböző környezeti tényezőkhöz való alkalmazkodás az eltérő élőhelyeken összességében nagyobb diverzitást eredményez – a változatosság pedig adaptív szempontból is potenciálisan fontos.

5.1.6. Szelektált gének és variációk házinyúl fajtákban

A szelektált régiókat megvizsgáltam erősen differenciált genetikai változatokat keresve is. Összesen 253 variáns rendelkezett 0.9-nél nagyobb Fst értékkel, 983 pedig 0.75-nél nagyobbval a házinyúl és üregi nyúl populációk között (5. táblázat, M3. Melléklet). Ebből a 983 változathoz 683 (69,48%) adódott teljesen fixáltnak a házinyulakban, míg csak 8 az üregi nyulakban. Ez egybevégt a házinyúlban a domesztikációs folyamat következtében várt alacsonyabb genetikai diverzitással is. A vizsgált változatok közül összesen 2 volt eltérő allélre fixált a házinyúlban és üregi nyúlban, mindkettő az ENSOCUG00000008635 génben, amely a humán és egér MAN1C1 gének ortológja (Ensembl), minden más változat polimorfizmust mutatott legalább az egyik populációban. A GeneCards (Stelzer *et al.*, 2016, <https://www.genecards.org/>) információja szerint testmérettel, valamint vérlemezké, és vér-fehérje mennyiséggel kapcsolatos. A gén fehérje terméke a Golgi-apparátusba és onnan az endoplazmatikus retikulumba történő transzportba történik.

Az erősen differenciált változatok közel fele ($473/983 > Fst 0,75$, $142/253 > Fst 0,9$) génekben található, de ezek közül összesen csupán 9 adódott exonokban, és 15 prediktált TFBS-ben. Gén szinten vizsgálva a legtöbb erősen differenciált változat a CFLAR génben található: összesen $61 > Fst 0,9$ és $132 > Fst 0,75$ genetikai változat. A második legtöbb az ENSOCUG000000038656 azonosítójú génben fordul elő: $40 > Fst 0,9$ és $131 > Fst 0,75$. Az Ensembl adatbázisa szerint ez az lncRNA gén nem rendelkezik ismert funkcióval, vagy ortológgal. A CFLAR domesztikációban betöltött feltételezhető szerepéről sem áll rendelkezésre sok információ, bár Dai *et al.* (2022) rávilágít a T-sejtes immunválaszban játszott szerepére madárinfluenza fertőzés során.

2. táblázat – Génekkal átfedő, feltételezetten szelektált régiók és azok géntartalma

Szelektált régió			Átfedő gének	
Kromoszóma	Start pozíció	Vég pozíció	Ensembl gén azonosító	Gén
2	8410001	8415001	ENSOCUG00000006802	LCORL
2	8420001	8435001	ENSOCUG00000006802	LCORL
2	8445001	8450001	ENSOCUG00000006802	LCORL
2	8510001	8515001	ENSOCUG00000006802	LCORL
2	8525001	8530001	ENSOCUG00000006802	LCORL
2	97360001	97365001	ENSOCUG00000006058	POLR1B
2	97370001	97376882	ENSOCUG00000006058	POLR1B
4	68490001	68510001	ENSOCUG00000008219	POC1B
4	68725001	68750001	ENSOCUG00000016507	ATP2B1
7	105410001	105425001	ENSOCUG00000034812, ENSOCUG00000007266	protein_coding, SCN1A
7	105935001	105955001	ENSOCUG00000017729	SCN7A
7	105695001	105715001	ENSOCUG00000005623	SCN9A
7	105740001	105785001	ENSOCUG00000005623	SCN9A
7	105875001	105915001	ENSOCUG00000017729	SCN7A
7	141860001	141930001	ENSOCUG00000011072	CFLAR
12	83310001	83390001	ENSOCUG00000034993, ENSOCUG00000032144	lncRNA, lncRNA
13	128055001	128068945	ENSOCUG00000008635	protein_coding
13	128590001	128598982	ENSOCUG00000034322	lncRNA
13	139035001	139109705	ENSOCUG00000016927, ENSOCUG00000016938, ENSOCUG00000016916	SH3BGRL3, UBXN11, CEP85
14	2235001	2245001	ENSOCUG00000039398	protein_coding
14	2445001	2450442	ENSOCUG00000015875	RFTN1
14	29940001	29965001	ENSOCUG00000001843, ENSOCUG00000001839	MSL2, PPP2R3A
14	77753269	77823271	ENSOCUG00000038656	lncRNA
14	83370001	83383479	ENSOCUG00000005429	BCL6
15	12905001	12925001	ENSOCUG00000002636	FAM160A1
16	34665001	34675001	ENSOCUG00000010372	MAP3K21
16	34690001	34695001	ENSOCUG00000010372	MAP3K21
16	34965001	34980001	ENSOCUG00000014363	PCNX2
16	34940001	34955001	ENSOCUG00000014363	PCNX2
16	35630001	35635001	ENSOCUG00000010297	SIPA1L2
16	35650001	35655001	ENSOCUG00000010297	SIPA1L2
18	63109193	63115001	ENSOCUG00000005958	ATRNL1

3. táblázat - Génekkal nem átfedő, de szelektáltak adódott genomi régiók

Kromoszóma	Start pozíció	Végpozíció
2	9520001	9530770
4	69050001	69085001
7	105585001	105605001
7	105840001	105865001
7	134021138	134040001
8	19610001	19623583
8	19980001	20000001
8	20030001	20040001
9	107735001	107755001
10	6185001	6190001
14	148705001	148720001
18	26680001	26700001
20	6305001	6325001
20	6281414	6300001

4. táblázat - Szelektált régiókban azonosított, szignifikánsan dúsultnak adódó transzkripció faktor kötőhelyek. Az első oszlop a genomon azonosított teljes mennyiséget, míg a harmadik az ebből szelektált régiókban található mutatója.

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

Számos más házasított fajhoz hasonlóan (Bai *et al.*, 2025; Krutikova *et al.*, 2024; Makvandi-Nejad *et al.*, 2012; Saif *et al.*, 2020; Wang *et al.*, 2022, stb.) szelektáltak adódott a testmérettel asszociált LCORL gén is.

A munkám során szelektáltként azonosított gének sokszor nem azonosak a más házasított fajokban találhatók, ugyanakkor gyakran kapcsolódnak a domesztikáció során általánosan érintett folyamatokhoz. Ilyenek például az immunrendszer működésében és az immunválaszban szerepet játszó gének - BCL6 (Liu, C. *et al.*, 2021), CFLAR (Dai, M. *et al.*, 2022; Zhang, H. *et al.*, 2025), RFTLN1 (Stelzer *et al.*, 2016) -, az idegrendszeri folyamatokban, jelátvitelben - SCN7A (Ke *et al.*, 2012; Stelzer *et al.*, 2016), SCN1A (Sawyer *et al.*, 2016), SCN9A (Singh *et al.*, 2009) - vagy az idegrendszeri és egyedfejlődésben - MSL2, POLR1B (Stelzer *et al.*, 2016) - érintett gének.

5. táblázat - A szelektáltnak adódott génekben található, házi- és vad üregi nyúl között erősen differenciált variációk száma. A szám a teljes gén régióban található - exon, intron, UTR - változatok összességét figyelembe veszi.

Gene ID	Per-Site Fst		Name
	>0.75	>0.9	
ENSOCUG00000001839	72	22	<i>PPP2R3A</i>
ENSOCUG00000001843	72	22	<i>MSL2</i>
ENSOCUG00000002636	9	0	<i>FHIP1A</i>
ENSOCUG00000005429	9	0	<i>BCL6</i>
ENSOCUG00000005623	39	0	<i>SCN9A</i>
ENSOCUG00000005958	2	1	<i>ATRNL1</i>
ENSOCUG00000006058	1	2	<i>POLR1B</i>
ENSOCUG00000006802	11	0	<i>LCORL</i>
ENSOCUG00000007266	3	0	<i>SCN1A</i>
ENSOCUG00000008219	24	5	<i>POC1B</i>
ENSOCUG00000008635	36	19	Novel (humán MAN1C1 ortológ)
ENSOCUG00000010297	11	0	<i>SIPA1L2</i>
ENSOCUG00000010372	8	0	<i>MAP3K21</i>
ENSOCUG00000011072	132	61	<i>CFLAR</i>
ENSOCUG00000014363	40	0	<i>PCNX2</i>
ENSOCUG00000015875	16	8	<i>RFTN1</i>
ENSOCUG00000016507	11	3	<i>ATP2B1</i>
ENSOCUG00000016916	58	38	<i>CEP85</i>
ENSOCUG00000016927	58	38	<i>SH3BGRL3</i>
ENSOCUG00000016938	58	38	<i>UBXN11</i>
ENSOCUG00000017729	42	0	<i>SCN7A</i>
ENSOCUG00000032144	97	33	novel, lncRNA
ENSOCUG00000034322	11	5	novel, lncRNA
ENSOCUG00000034812	3	0	novel
ENSOCUG00000034993	97	33	novel, lncRNA
ENSOCUG00000038656	131	40	novel, lncRNA
ENSOCUG00000039398	5	1	novel

Az immunválaszban és jelátvitelben szerepet játszó gének közé tartoznak a 9 exonikus variáció által érintett *BCL6*, *CFLAR* és *SCN7A*. Exonikus variáció található még az ENSOCUG00000034993 azonosítójú lncRNA génben is.

Az SNPEff (Cingolani *et al.*, 2012) segítségével végzett funkcionális annotáció alapján szinte az összes variáció kis hatásúnak adódott. Ez az eredmény egybevág más, domesztikációval kapcsolatos eredményekkel, amelyek alapján nem nagy hatású, vagy új variációk, hanem számos kis hatású variáció frekvenciájának eltolódása vezet a házasítás során megjelenő változásokhoz (Andersson és Purugganan, 2022; Carneiro *et al.*, 2014b; Zheng *et al.*, 2020).

Számos üregi nyúl gén funkciója nem jól, vagy egyáltalán nem ismert, és az OryCun2.0 referenciagenom is sokkal kevésbé alaposan feltérképezett, mint a humán, vagy akár az egér

genom. A gének annotációját a genom töredezettsége, és egyes szakaszok problémás annotációja is nehezíti. Az UTR szakaszok *de novo* annotációja például a mai napig problémás, a fehérjét nem kódoló, lncRNS gének pedig történelmi okokból gyakorlatilag minden organizmusban kevésbé ismertek – sokáig nem tekintették ezeket fontosnak.

Teljesebb, az ún. T2T (telomer-to-telomer) szinthez közelítő genomok használata könnyebbé teszi a genom szintű vizsgálatokat. Annak ellenére, hogy a genom teljes annotációja így is jelentős kihívás, a nagy folytonosságú, jó minőségű szekvenciának köszönhetően biztosabb a gének teljes, nem töredezett jelenléte, és a lehető legtöbb szekvenciát kromoszómán elhelyezve a kapcsoltsági viszonyok is átláthatóbbá válnak.

5.2. Új havasi nyúl (*Lepus timidus*) referenciagenom összeszerelése

5.2.1. Szekvenálási eredmények és genom összeszerelés

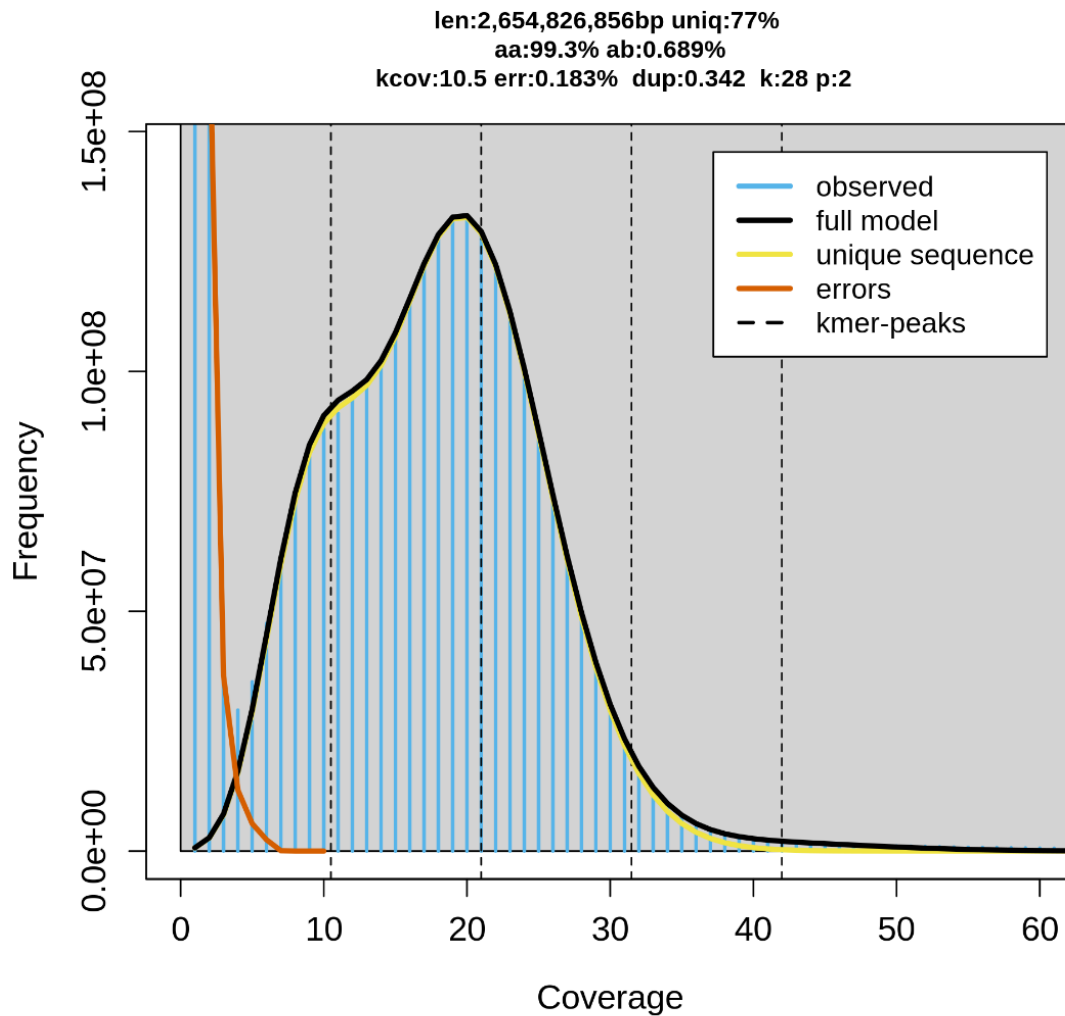
A havasi nyúl genomjának összeszerelésénél fibroblaszt sejtkultúrákból izolált DNS-t használtunk a PacBio HiFi szekvenáláshoz, a csoportunk által korábban összeszerelt mezei nyúl genomhoz hasonlóan (Michell *et al.*, 2024). A módszer előnye, hogy a sejtkultúrákból könnyebb nagy mennyiségű intakt, nagy molekulású DNS-t kinyerni, amely elengedhetetlen a referencia minőségű genom összeszereléséhez megfelelő szekvenáláshoz. A genom összeszerelése során nem mutatkoztak olyan genomi átrendeződések vagy aneuploidiák, amelyek a sejtek kultúrában tartására lettek volna visszavezethetők. Bár ilyen változások előfordulhatnak, ezek nagyobb eséllyel történnek hibás DNS javítási- és osztódási mechanizmusú rák-sejtvonalakban. Ugyanakkor a sejtvonalak hosszú távú fenntartása kerülendő, és az analízishez alacsony passzázs szám preferált, mivel idővel előfordulnak genetikai változások, és ezek elterjedhetnek a sejtek áttöltésakor (Wenger, S. L. *et al.*, 2004).

A PacBio HiFi szekvenálásból származó leolvasások hossza $N50 = 19,97$ kbp-nek adódott. Az $N50$ a legrövidebb read hosszát jelenti a teljes szekvencia hosszúság 50%-ánál, amennyiben a readokat a leghosszabbtól a legrövidebb felé méret szerinti sorba rendezzük. A becsült lefedettség 21-szeresnek, a genom becsült hosszúsága k-mer elemzés alapján 2,65 Gbp-nek adódott (**9. ábra**). Az adatok minősége és mennyisége elegendő volt az összeszerelés során haplotípusok képzéséhez.

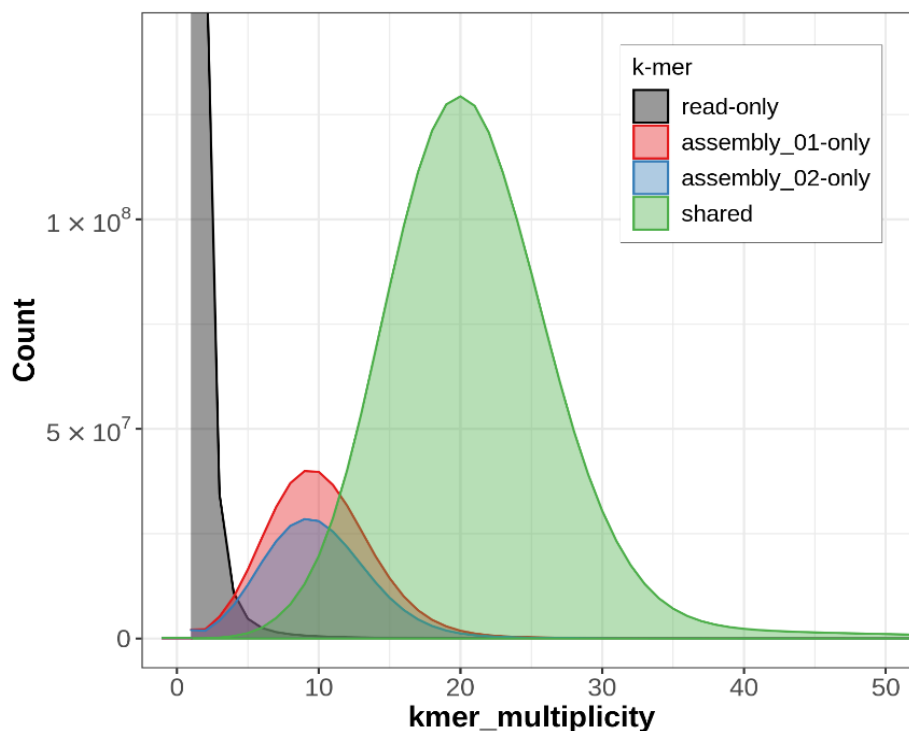
Az Illumina Hi-C szekvenálás 995 011 507 leolvasás párt (150bp x 2) eredményezett, amely a genom kb. 112-szeres lefedettségét jelenti.

A HiFiasm összeszerelési lépés egy 2 838 807 612 bp hosszúságú, 2332 contigból álló elsődleges, és egy 2 654 293 704 bp hosszúságú alternatív összeszerelést eredményezett. A leghosszabb contigok 22 és 26,6 Mbp hosszúságúak voltak, az $N50$ pedig 4,7 és 4,8 Mbp. Az elsődleges összeszerelés a mercury (Rhie *et al.*, 2020b) k-mer spektrumot felhasználva, annak teljessége alapján választottam ki (**10. ábra**). A Hi-C adatok felhasználásával javítottuk a téves összeszereléseket, és scaffoldokká szereltük össze a contigokat. A scaffolding és deduplikáció után az elsődleges szekvencia ezt követően 2 654 826 856 bp hosszúságú volt, a leghosszabb scaffold 181 Mbp, míg a scaffold $N50 = 127$ Mbp.

GenomeScope Profile



9. ábra – GenomeScope (Ranallo-Benavidez et al. 2020) generált k -mer profil. A vártak megfelelő diploid profil, amely a k -mer lefedettséget (Coverage) és frekvenciát (Frequency) mutatja. A 21-szeres és 10,5-szeres (kcov) lefedettségénél láthatók a homozigóta és heterozigóta csúcsok. Homozigóta szekvenciák (becsült aa: 99,3% a teljes szekvenálásban) esetében adott k -mer várhatóan kétszeres lefedettséget mutat a heterozigótákhoz (becsült ab: 0,689%) képest. A spektrumon további csúcsok nem jelennek meg, ez alapján duplikációk vagy aneuploidiák nem várhatók. A szekvencia teljes várt hossza (len) 2 654 826 856 bp, ennek 77%-a (uniq) egyedi k -mer volt. A duplikált (dup) és szekvenálási hibából adódó (err) k -mer mennyiség 0,183% és 0,342%-nak adódott. A beállított k -mer hosszúság (k) 28 volt, a feltételezett ploidia (p) pedig 2. (Fekete et al., 2025a)



10. ábra – Merqury (Rhie et al., 2020b) generált k-mer spektrum az első genomösszeszerelési lépést követően. A piros és kék csúcsok jelzik azokat a k-mereket – nagyrészt heterozigóta lókuszok – amelyek csak az egyik, vagy a másik contig készletben vannak jelen. Az elsődlegesként kiválasztott a teljesebbnek feltételezhető, több szekvenciát tartalmazó (piros, assembly_01) összeszerelés volt. A „k-mer multiplicity” statisztika megegyezik a GenomeScope által meghatározott lefedettséggel. (Fekete et al., 2025a)

5.2.2. Az összeszerelt genom

Az összeszerelt genom a vártnak megfelelően 23 autoszómát, valamint X és Y kromoszómát tartalmaz, és megfelel az az Earth Biogenome Project (EBP) által ajánlott 6.C.Q40 minőségi kritériumnak (Lawniczak et al., 2022). A kritériumok között szerepel az N50 > 1 Mbp szintű kontinuitás, a kromoszóma szintű összeszereltség, és az 1/1000-nél alacsonyabb nukleotid hiba arány is.

A *Lepus timidus* esetében irodalmi adatok alapján a várt haploid genom méret 3,28 Gbp (Vinogradov, 1998), amely 23 autoszómát, valamint az X és Y ivari kromoszómákat tartalmazza. Ezzel szemben saját PacBio HiFi szekvenálási adataink k-mer elemzése alapján összeszerelés előtt a genom várt mérete 2,65 Gbp-nek adódott, míg a végleges méret – összeszerelés, dekontamináció és kézi kuráció után – 2,70 Gbp. Az NCBI Assembly adatbázisában korábban megtalálható *Lepus timidus* referencia (GenBank accession GCA_040893245.2, Marques et al., 2020) hossza szintén 2,70 Gbp, igen közel saját eredményeinkhez, szemben Vinogradov közel 500 Mbp-vel hosszabb becslésénél.

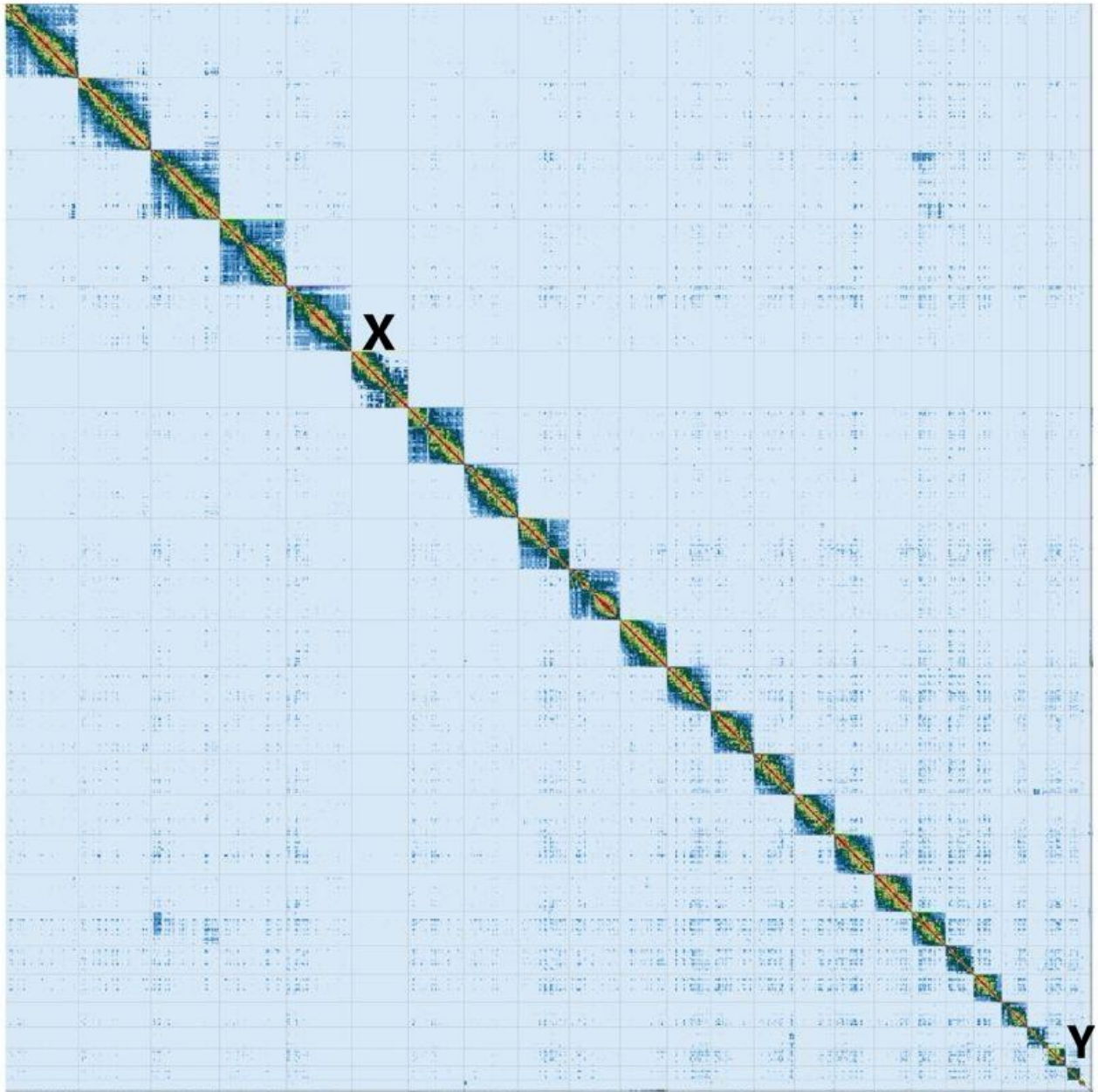
A meglehetősen nagy eltérést Vinogradov és saját eredményeink között több tényező együttes befolyása okozhatja. Vinogradov kalkulációja áramlási citometriás mérésekre, és egy ismert genom hosszúságú fajra alapul. A referenciaként szolgáló faj a havasi nyúltól rendszertanilag távol álló gyepi béka (*Rana temporaria*) volt. A kromoszómák eltérő szerveződése is befolyásolhatja a kalkuláció eredményét. A flow-citometriás, méret alapján történő becslésnél a nagy pontosságú leolvasások alapján történő összeszerelés összességében pontosabb becslést tesz lehetővé.

A szekvenálási adatok alapján történő genom összeszerelésnél előfordulhatnak pontatlanságok, főleg a nehezen összeszerelhető repetitív szakaszok környékén, illetve erősen heterozigóta szakaszok esetében, ahol a két haplotípust a nagy mértékű eltérés miatt az algoritmus egy haplotípusba tartozó, különböző szakaszokként értelmezi. Ez utóbbit nevezik haplotipikus duplikációnak, az ilyen szakaszokat pedig haplotignak. Nagy részük a deduplikáció során a `purge_dups` (Guan, D. *et al.*, 2020) szoftverrel automatikusan eltávolítható, a megmaradó problémás szakaszok a kézi kuráció során általában jól kiszűrhetőek. A Hi-C szekvenálási eredményeket mutató kapcsolódási térképen („contact map”), főleg hosszabb szakaszok esetében, jól láthatóak és könnyen eltávolíthatóak a végső genomból. Nehezebben megoldható problémát jelentenek az erősen repetitív szakaszok. Ezeket az összeszerelő algoritmusok gyakran lerövidítik („collapse”). Minél hosszabb egy ilyen szakasz, annál nehezebben értelmezhető az algoritmusok számára, és kézi kiigazításuk is nehézségekbe ütközik: a repetitív régiók jellemzője az erős Hi-C szignál, és sok esetben a számos apró, töredezett contig, amelyek tényleges sorrendje nehezen, vagy egyáltalán nem megállapítható a genomon. A *Lepus timidus* referencia Hi-C kontakt térképét a kézi kuráció után az **11. ábra** mutatja.

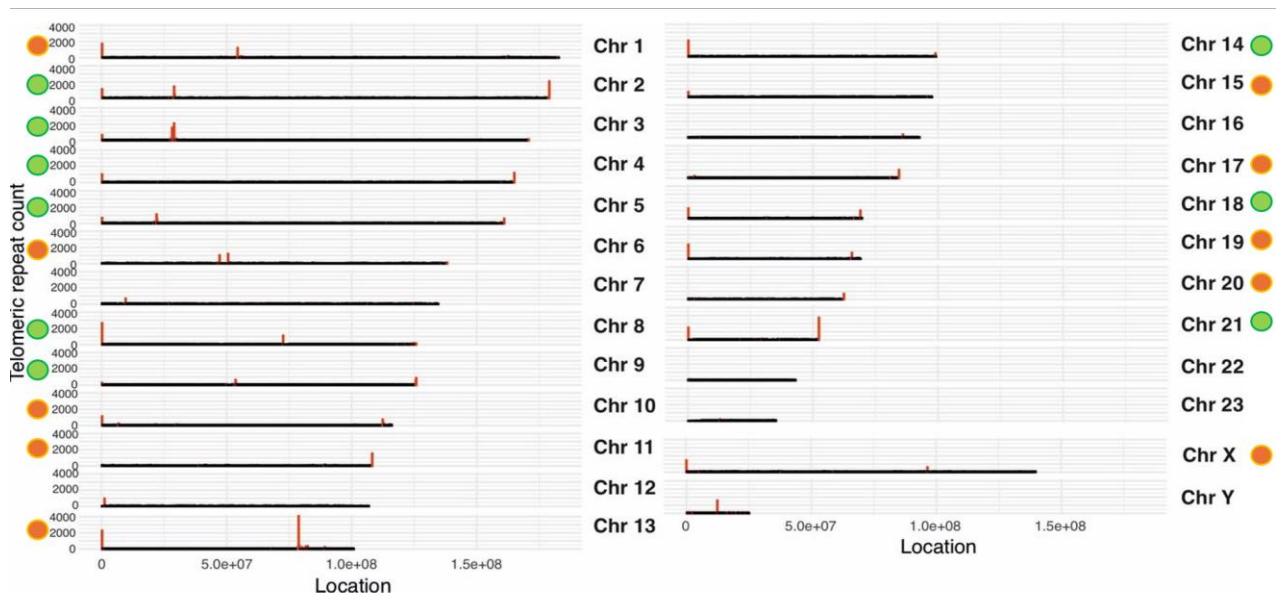
A genom repetitív szakaszainak térképezése során a RepeatModeler (Flynn *et al.*, 2020) segítségével 28 386 egyedi repetitív elemet találtam, amelyet a program 809 családba sorolt be. A genom repetitívként megjelölt szakaszai összesen a teljes hossz 42,35%-át teszik ki, amely meglehetősen magas. Ugyanakkor ez az eredmény nagyon hasonló a mezei nyúl genomjának összeszerelése során tapasztaltnak. Ez utóbbiban a teljes genom hosszának 46,8%-a adódott repetitív szekvenciának (Michell *et al.*, 2024). Ezek az értékek mindkét genom esetében nagyjából kétszeresét teszik ki a k-mer alapú előzetes becslésnek. A becsült érték 23,0% volt a havasi, és 26,7% a mezei nyúl esetében. Bár az újonnan összeszerelt *Lepus* genomokban az értékek nagyon hasonlóak, a korábbi, írországi mintából (*Lepus timidus hibernicus*) készült havasi nyúl draft genomban a genom összesen 28%-a lett repetitív elemként maszkolva (Marques *et al.*, 2020). Ennek legvalószínűbb oka, hogy Marques és munkatársai (2020) Illumina rövid leolvasásos szekvenálásokra alapozták munkájukat, amely különösen megnehezíti a repetitív régiók összeszerelését.

Sikeresen azonosítottam telomer szekvenciákat 9 kromoszóma mindkét végén (**12. ábra**), amely ezeknek a kromoszómáknak teljességére utal. A telomer szakaszok repetitív mivolta miatt ezek sokszor még a nagy pontosságú hosszú reades szekvenálási módszereknek is kihívást jelentenek, hasonlóan a fentebb említett, egyéb repetitív szakaszokhoz képest. Így jelenlétük jó jelzés egy összeszerelt kromoszóma teljességét illetően, de hiányuk nem feltétlenül jelent komoly hiányosságot a szekvencia teljes hosszának tekintetében.

Megfigyeltem ezen felül 13 kromoszómán (1., 2., 3., 5., 6., 7., 8., 9., 10., 13., 20., X. és Y kromoszómák) telomer elemek jelenlétét más, nem a kromoszóma végén található pozíciókban (**12. ábra**). Az ilyen pozíciókban magas ismétlésszámmal előforduló telomer elemek jelenléte a korábbi, fluoreszcens in-situ hibridizációs (FISH) staining módszerrel kapott eredményeknek megfelelőek (Forsyth *et al.*, 2005). Ezek a régiók az evolúció során bekövetkezett transzlokációs események maradványai lehetnek, amelyek hozzájárultak a jelenlegi kariotípus kialakulásához.



11. ábra - Lepus timidus Hi-C kontakt térkép a kézi kurációt követően. A térkép a 23 autoszómát, valamint az X és Y kromoszómákat is tartalmazza. Az X és Y kromoszóma meghatározása lefedettség, és az azokon található gének alapján történik. A térképen vizuálisan is jól látható az X kromoszóma: a feleakkora lefedettség az autoszómákéhoz képest látványosan kevesebb szórványos szignált eredményez a genom többi része és az X kromoszóma között. Az autoszómák számozása méret szerint történt, azaz a bal felsőtől a jobb alsó sarok felé. (Fekete et al., 2025a)



12. 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).

5.2.3. A genom minőségi ellenőrzése

Az összeszerelt genomon lefuttatott BUSCO (Manni *et al.*, 2021) teszt eredményei alapján az közelítőleg teljes. A BUSCO gének 1:1 ortológjainak jelenléte alapján segít képet alkotni a genom teljességéről.

A havasi nyúl genom teljességének teszteléséhez két adatbázist, a mammalia_odb10 emlős, és a glires_odb10 leporid ortológ génkészletet használtam fel. A mammalia_odb10 adatbázisban található ortológok közül 95.1% teljesen (egy kópiában: 92,3%, duplikáltan: 2,7%), 0.8% fragmentáltan van jelen, és 5,6% hiányzik. A teszthez a glires_odb10 adatbázist felhasználva százalékosan kevesebb (93,2%) ortológot találtam meg teljes egészében, ebből 90,6% egy kópiában, 2,7% pedig duplikáltan volt jelen. Ebben az esetben az ortológok 1,1% -a volt jelen töredékesen, és 5,6% hiányzott teljesen. Ugyanakkor az emlős adatbázis 9226, míg a leporid adatbázis 13798 BUSCO ortológ csoportot tartalmaz, emiatt a találatok nagyobb száma ellenére százalékosan kevesebb a találat. A BUSCO eredmények részleteit az [M7. Melléklet](#) foglalja össze.

A kézi kuráció a scaffoldok mennyiségét 103-ról 85-re csökkentette az elsődleges, és 128-ról 109-re az alternatív összeszerelésben. Ezt követően a Hi-C readok újratérképezésével vizuálisan ellenőrizhető, hogy a Hi-C kapcsolatok valóban támogatják-e a genom feltételezett szerveződését. Amennyiben szükséges, ezek a lépések többször ismételtethők.

A végső genomban a szekvenciák 99,58%-a autoszómához vagy alloszómához rendelhető volt (**6. táblázat**). A genom statisztikai jellemzőit a Quast (Mikheenko *et al.*, 2018) és gfastats (Formenti *et al.*, 2022) programok eredményei alapján a [M7. Melléklet](#) foglalja össze.

A genom publikálása előtti, az NCBI által végzett utolsó kontamináció szűrés a korábbi két szűrés (általam és a Wellcome Sanger Institute kurátorai által) ellenére T-antigén vektor inszerciókat fedezett fel a 17. és 19. kromoszómán. A szekvenciákat manuálisan eltávolítottam és azonos

hosszúságú „N” szekvenciával helyettesíttem, mivel ismeretlen, hogy milyen hosszúságú és szekvenciájú szakasz került esetleg törlésre a beépülés során. A vektorok jelenléte emiatt annak ellenére, hogy feltehetően nem jelent komoly problémát, nem ideális. Az intergénikus szakaszban talált vektor kontamináció szekvencia azonosság alapján a fibroblaszt sejtvonalt immortalizációjához használt rekombináns expressziós vektor (Gaertner *et al.*, 2023) volt. Az immortalizáció, bár nem feltétlenül szükséges a dolgozatomhoz végzett munkához, megkönnyíti a sejtvonalak kezelését és minimalizálja a fenotipikus variációk jelenlétét, amely előnyös a sejtvonalt további kísérletekben való felhasználásánál (Fekete *et al.*, 2025a), és szükség esetén megkönnyíti az ugyanabból az egyedből történő új szekvenáláshoz való mintavételt is. Ezzel együtt referenciagenomok összeszerelésénél primér sejt kultúrák használata ajánlott, hasonló kontamináció elkerülése érdekében.

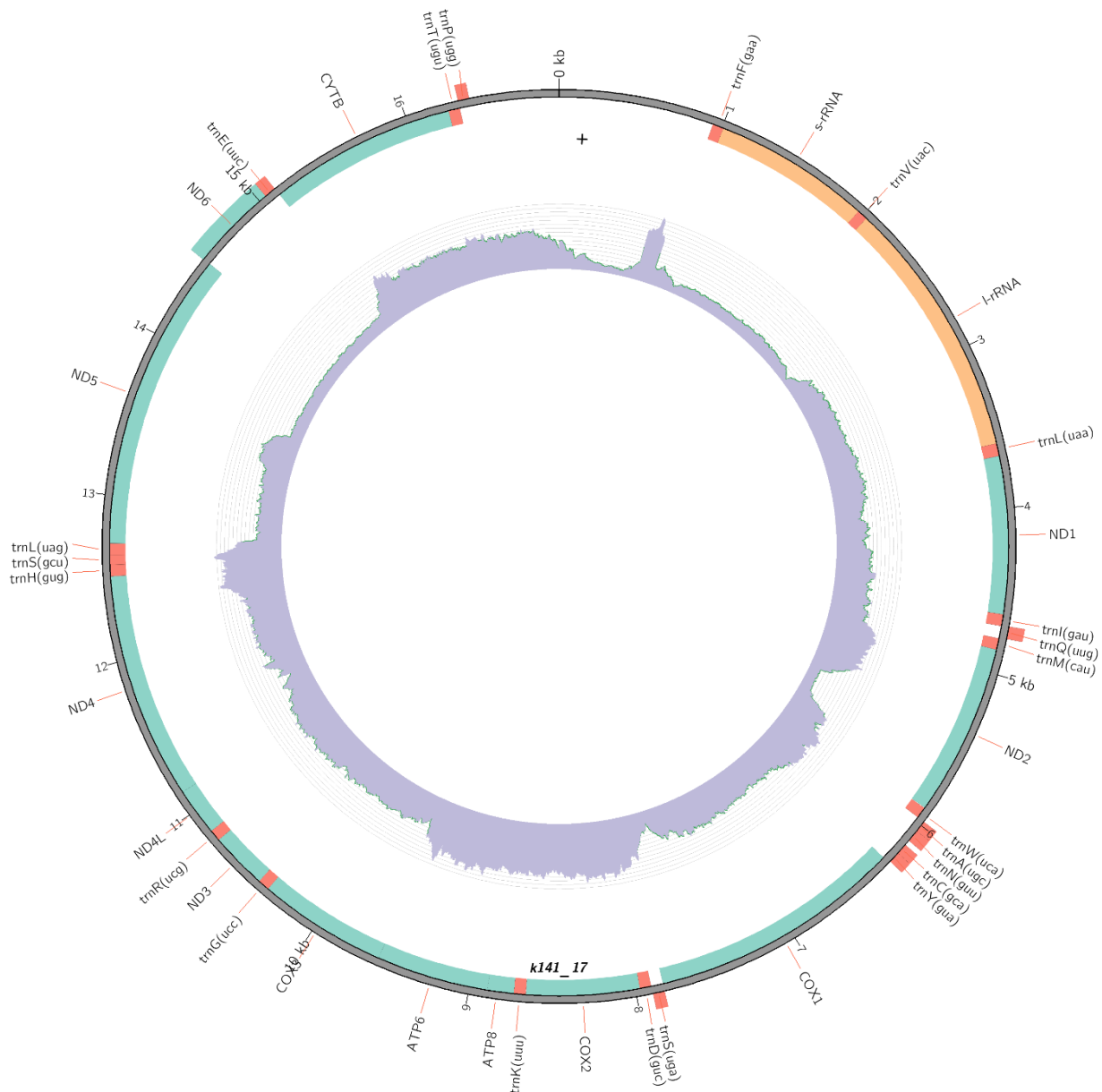
6. táblázat - Összefoglaló táblázat az összeszerelt genom tulajdonságairól – A szekvenciák kromoszómához és scaffoldokhoz rendeltsége. Az „autoszómához kapcsolt” olyan scaffoldokat jelent, amelyek autoszómához tartozók, de azon belüli elhelyezkedését nem tudtuk meghatározni.

	Hossz [bp]	Teljes hossz %-a	Mennyiség
Autoszómák	2,519,453,057	93.48	23
Autoszómához kapcsolt	1,279,896	0.05	1
Egyéb scaffold	11,202,567	0.42	58
X	139,721,860	5.18	1
Y	24,927,870	0.92	1
MT	17,482		1
Összes	2,695,305,354		85

Érdekes módon a szintén csoportunk által összeszerelt mezei nyúl referenciagenomban (Michell *et al.*, 2024a) nem volt jelen hasonló kontamináció, amely nagy valószínűséggel arra utal, hogy a sejtvonalt immortalizációja sikertelen volt. Az pedig, hogy még a harmadik kontamináció szűrési lépésnél is jelent volt – habár minimális – idegen eredetű szekvencia, rámutat a többszörös ellenőrzés szükségességére egy genom publikációja előtt.

5.2.4. A mitokondriális szekvencia

Az összeszerelt mitokondriális szekvencia (**13. ábra**) hossza az LT1 sejtvonalt esetében, amelyből a referencia készült, 17 482 bp. Érdemes azonban a referencia felhasználásánál figyelembe venni, hogy a havasi és mezei nyulak mitokondriális DNS-ére egyaránt jellemző a nem kódoló szakaszban jelenlévő eltérő mennyiségű hosszú ismétlődő szekvencia, amelyek a long repeat (LR) régiót alkotják. Az LR régió fajok közti variabilitása korábban is ismert volt (Lunt *et al.*, 1998), *Lepus* fajokban azonban egyedek között is változónak bizonyult: finnországi havasi nyulak esetében a leggyakoribb ismétlés szám 5 (49%), de nem sokkal marad el tőle a 6 (40%) sem. Emellett 4-es és 7-es ismétlés szám, valamint ezekre heteroplazmikus egyedek is előfordulnak (Tapanainen *et al.*, 2024). A repetitív szakasz hosszának jelentősége még nem tisztázott.



13. ábra - Az LT1 sejtvonal összeszerelt, annotált mitokondriális szekvenciája. A belső (lila) kör mutatja a szekvenálási lefedettséget. A szakaszokban mutatkozó ugrásszerű eltérések az Illumina szekvenálást megelőző amplifikációs lépés következményei. A külső kör mutatja a gének elhelyezkedését – a géneket nem tartalmazó szakasz az LR régió.

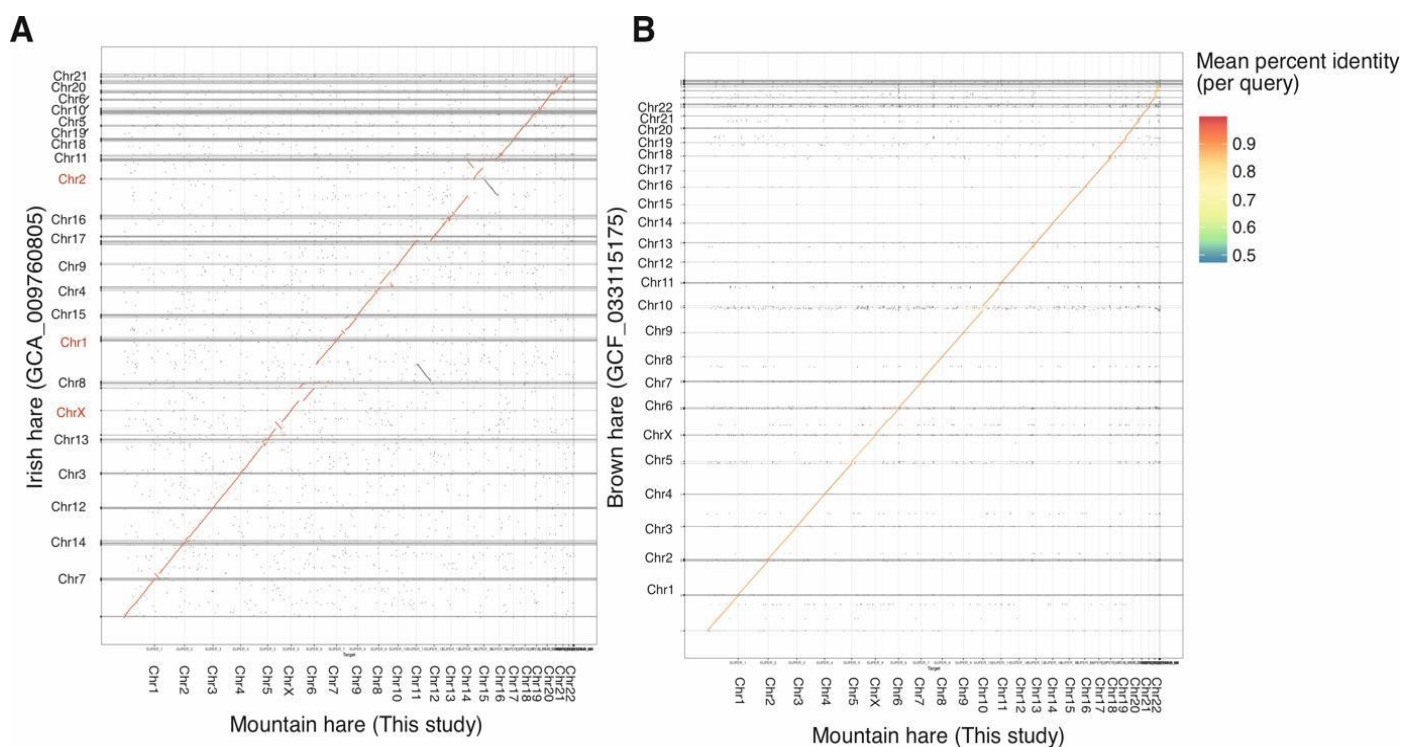
5.2.5. Az adatok elérhetősége

Az összeszerelt genom (GenBank: GCA_040893245.2) és a hozzá tartozó szekvenálási adatok (Azonosítók: BioSample SAMN41430840 és BioProject PRJNA1112569 az elsődleges haplotípushoz, PRJNA1112568 az alternatívhoz) nyilvánosan elérhetők az NCBI oldalán.

5.2.6. Az új *Lepus timidus* referencia összevetése korábban összeszerelt *Lepus* genomokkal

A Minimap2 (H. Li, 2018) hosszú leolvasás illesztő program az ír havasi nyúl (*Lepus timidus hibernicus*) genom contig szekvenciáinak 99,95%-át volt képes illeszteni az általunk összeszerelt genomra, ennek 97,57%-át „primary alignment”-ként. A „primary alignment”, azaz elsődleges illesztés az illesztésnél a legjobb pontszámmal illeszkedő szekvenciát jelenti. A két genom nagyon hasonlóknak bizonyult szekvenciájában.

Az összeszerelt genomok szerveződésében a szintén erős hasonlóság mellett jellemző néhány fő különbség. Az ír havasi nyúl genomjának összeszereléséhez az üregi nyúl (*Oryctolagus cuniculus*) kromoszóma szerveződését (Beklemisheva *et al.*, 2011) használták fel (Marques *et al.*, 2020). Ennek hatása jól látszik nem csak az általunk összeszerelt havasi, hanem a mezei nyúl genommal való összehasonlításban is. A legfőbb különbség, hogy a 7-es és 12-es *Lepus* kromoszómák együtt az üregi nyúl 1-es kromoszómájának felelnek meg, míg a 13-as és 16-os kromoszómák az üregi nyúl 2-es kromoszómájának (14. ábra). Ez megfelel a havasi és üregi nyúl közötti kariotípus különbségnek is: havasi nyúl $2n = 48$, míg üregi nyúl $2n = 44$.



14. 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, amelyet a világosabb, sárgába hajó szín is mutat, és – a vártaknak megfelelően – kevés a kromoszóma-átrendezés. (Fekete *et al.*, 2025a)

Erre vezethető vissza az is, hogy bár az ír havasi nyúl (*Lepus timidus hibernicus*) minta közelebbi rokonságban áll a finnországi havasi nyúllal (*Lepus timidus timidus*), az összeszerelt genomokat összehasonlítva mégis úgy tűnik, mintha a finnországi havasi nyúl a mezei nyúlhoz állna közelebb a kromoszómák szerveződését és szerkezetét illetően (14. ábra). A Minimap2 (H. Li, 2018) a mezei nyúl esetében a contigok 99,87%-át volt képes az általunk összeszerelt genomra illeszteni, amely valamivel kevesebb az ír havasi nyúl esetében megfigyeltnél. Ugyanakkor az elsődlegesen térképezett (primary mapped) contigok ebben az esetben csak az illesztések 78,77%-át teszik ki, amely több kimérikus, vagy részleges illeszkedés jelenlétére utal, kisebb különbségek jelenlétét mutatva a kromoszóma szekvencia szerveződésében.

A genomok hasonlósága jól mutatja a két faj evolúciós közelségét, és a vártaknak megfelel. A mezei nyúl és az üregi nyúl evolúciósan elég közel állnak ahhoz, hogy a két faj termékeny hibrid

utódokat tudjon létrehozni (Levänen *et al.*, 2018a; 2018b), de elég messze ahhoz, hogy bár Finnország területén aktív hibridizációs zóna létezik, a két faj ne mutassa az összeolvadás jeleit.

Az átlagos szekvencia hasonlóság, amelyet az illeszkedett contigok közötti pontosan illeszkedő bázisokból számítottam ki, a havasi nyúl és mezei nyúl összehasonlításban alacsonyabbnak adódott, mint a két havasi nyúl genom szekvencia között. Ez a vártnak megfelelő eredmény két rokon faj, illetve egy faj két alfaja közötti összehasonlításakor. Az átlagos hasonlóság értéke 96,77%-nak adódott a két havasi nyúl alfaj között, és 93,09%-nak a havasi nyúl és a mezei nyúl között (**14. ábra**). Az átlagos szekvencia hasonlóság számításához választott módszeremmel kapcsolatban érdemes megjegyezni, hogy azt kifejezetten a teljes genom összehasonlításához választottam. A kódoló szekvenciák összevetése ettől eltérő eredményt, és valószínűleg nagyobb hasonlóságot eredményezne, azok erős konzerváltsága miatt.

A *Lepus timidus* és *Lepus europaeus* genomon azonosított telomer szakaszok ismétlődő szekvenciái azonosnak adódtak ([AACCCT]_n), és mindkettőn egyaránt találhatóak telomer szekvenciák egyes kromoszómák belső részein is (Michell *et al.*, 2024).

5.3. Mitokondriális rekombináció előfordulásának megerősítése emlős sejtekben

5.3.1. Sejtvonalak és szekvenálás

A humán eredetű sejtvonalak mtDNS-e az afrikai L3 (HeLa) és az európai U5 (KSS) haplocsoportba tartozónak adódott. A heteroplazmikus, BH10 sejtvonalak jellemzője az U5 (KSS) haplotípuson található nagyméretű deléció. Továbbá a sejtvonalba jelenlévő restrikciós hely polimorfizmus miatt létrejövő fragmenthossz különbségek alapján a rekombináció nem csak szekvenálós vizsgálatokkal, hanem transzmissziós elektronmikroszkópia (TEM), vagy Southern blot segítségével is kimutatható (Fragkoulis *et al.*, 2024).

MGME1 knockout mivoltukból adódóan a sejtekben lineáris mtDNS fragmensek is nagy számban előfordulnak. Az MGME1 – a POLG mellett – a lineáris mtDNS lebontásáért felelős fő nukleáz (Peeva *et al.*, 2018). Az MGME1 mennyiségének csökkenése a lineáris mtDNS formák jelentős akkumulációjához vezet, és az mtDNS kettős száltörései esetében a javítás válik preferálttá a lebontással szemben. A rekombináció pedig a kettős száltörés javításának és a replikációs villa helyreállításnak a fő mechanizmusa. Az MGME1 knockout és sejtvonalak emiatt alkalmasak lineáris mtDNS-sel kapcsolatos megfigyelésekre, például replikációs folyamatok vizsgálatára (Torregrosa-Muñumer *et al.*, 2019). Megfigyelték, hogy az MGME1 mennyiségének csökkenése a sejtben előforduló cruciform mtDNS-ek mennyiségét is növeli. A mintákban maradó lineáris fragmentek leggyakrabban a replikációs origó környékéről származó abortált replikációs intermedierek (Kornblum *et al.*, 2013; Nicholls *et al.*, 2014; Peeva *et al.*, 2018; Torregrosa-Muñumer *et al.*, 2019).

A HP202B egér sejtvonala eredetileg körülbelül 20:80 NZB:Balb/c arányban volt heteroplazmikus. Az MGME1-KO klónok közül emiatt azok a kultúrák kerültek felhasználásra, amelyekben a heteroplazmikus aránya körülbelül 64:36 arányig eltolódott, ezzel növelve a rekombináció és annak megfigyelésének esélyét. A sejtvonala a Western blot vizsgálatok eredménye alapján csökkent MGME1 expressziót mutatott, annak teljes hiánya helyett, amely a knockout részleges sikerességére utal. Ezek a feltehetően heterozigóta sejtek is tartalmaztak azonban perzisztens lineáris DNS szekvenciákat (Fragkoulis *et al.*, 2024).

Két szekvenálási kísérlet történt, egy az Oxford Nanopore Technologies (ONT), egy pedig a Pacific Biosciences (PacBio) HiFi technológiájával. A Nanopore szekvenálás MinION eszközzel történt, időben megelőzve a Helsinki Egyetemen végzett PacBio HiFi szekvenálást. Mivel a

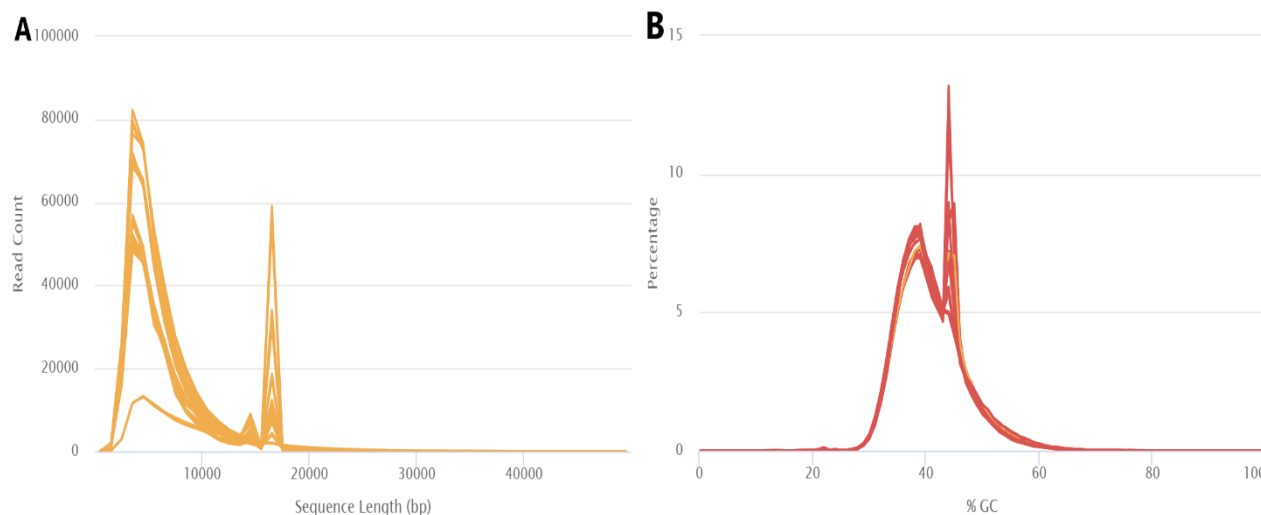
Nanopore technológiával nyert szekvenciák nem bizonyultak elég megbízhatónak a nagy pontosságú elemzéshez, amit a rekombináció molekulánkénti kimutatása igényelt, a szekvenálás megismétlésre került a második módszerrel – munkám során ennek a második szekvenálás eredményeire támaszkodva fejlesztettem ki a rekombináció detektálására alkalmas módszert. Bár a wt és Δ típusú mtDNS-t tartalmazó humán BH10 sejtvonal pont azért került kiválasztásra, hogy a deléció jelenléte segítse a rekombináció kimutatását, és ez a mikroszkópos és molekuláris biológiai laboratóriumi vizsgálatok során szükséges is volt, végül a szekvenálások esetében nem bizonyult elégséges különbségnek. Az egyes leolvasások és szakaszaik haplotípus-hovatartozásának megállapításában a legnagyobb szerepet az SNP-k kapták.

A PacBio HiFi technológiás szekvenálással az Illumina rövid leolvasásos szekvenáláséhoz hasonló szintű pontosságú adatok nyerhetők, de ebben az esetben a leolvasási hosszúság akár a teljes mitokondriális genom hosszúságát lefedheti.

Megjegyzendő, hogy az említett első kísérlet óta a Nanopore technológia sokat fejlődött, pontosságát és az elérhető maximális lefedettséget tekintve is. Ma már megfelelő pontosságú lenne a vizsgálathoz, és az általam alkalmazotthoz hasonló elven lehetséges lenne rekombináció felderítésére alkalmazni (Kolmogorov *et al.*, 2023).

5.3.2. A szekvenálás minőségi ellenőrzése és nem-mitokondriális DNS szűrése

A PacBio leolvasások alapvető minőségi szűrési lépései nagyon hasonlóak a referenciagenom összeszerelésnél alkalmazottakhoz. Ezek sokkal inkább a technológián, mint a felhasználási célon alapuló minőségbiztosítási lépések. A szekvenálás összességében kiváló minőségűnek bizonyult (<0.01% rossz minőségű, vagy adaptert tartalmazó leolvasás) ugyanakkor megfigyelhetőek voltak anomáliák, amely más esetben problémára utalnának, esetünkben viszont jól mutatják a mitokondriális szekvenciák nagy arányú jelenlétét az összes szekvenálási adat között (**15. ábra**).



15. 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.

Ebben az esetben szükség volt egy további szűrési lépésre is, annak biztosítására, hogy a további munkához csak tisztán mitokondriális eredetű szekvenciákat használják fel. A referenciagenom összeszerelés esetében nem volt hasonló probléma: a referenciagenomnak lehetőség szerint a

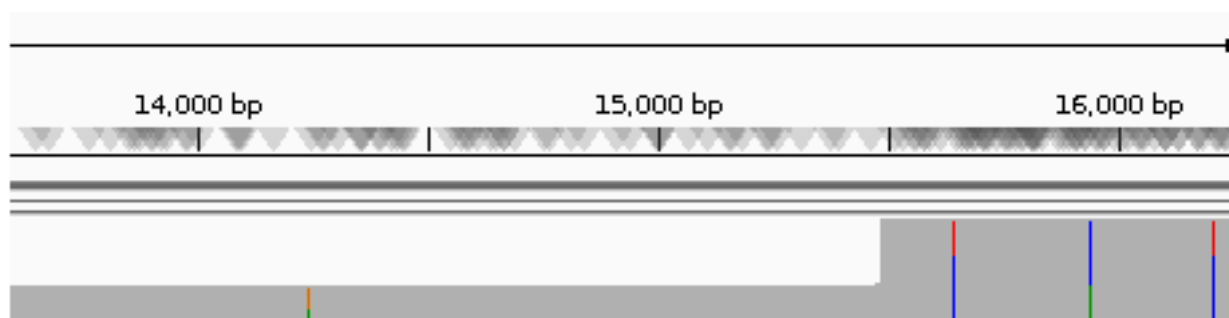
minta organizmus teljes DNS állományát tartalmaznia kell. A rekombináció vizsgálatánál csak specifikus szekvenciákra voltunk kíváncsiak.

Az mtDNS-re való szűrés azért is fontos lépés volt, mert a két felhasznált sejtvonal közül csak a BH10-ből került előzetesen tisztításra az mtDNS, míg a HP202B sejtvonalból teljes DNS szekvenálás történt – a referenciagenom készítésénél alkalmazotthoz hasonlóan. Ezen felül a tisztítás ellenére a BH10 szekvenálási adatok is nagy mennyiségű nukleáris szekvenciát tartalmaztak, amelyet a **15. ábra** nem-mitokondriális csúcsai is jól mutatnak.

5.3.3. Haplotípus különbségek - megbízható változatok

A kísérletek korai szakaszában felmerült, hogy a Δ mtDNS eliminálódhatott a BH10 sejtvonalból. Ezt a laboratóriumi vizsgálatok (Fragkoulis *et al.*, 2024) mellett a pbsv v2.9.0 (<https://github.com/PacificBiosciences/pbsv>) strukturális variáció hívó szoftver által a megfelelő pozíción, megfelelő hosszúsággal heterozigótaként hívott deléció jelenléte egyértelműen megcáfolta. Bár valóban valamivel kevesebb Δ típusú szekvencia volt jelen, a két eltérő haplotípus jelenlétét diploidként kezelve ez elegendő volt ahhoz, hogy a szoftver heterozigóta variációként azonosítsa azt. Érdekes módon, bár a sejtvonalak megfelelő szelektálása segítségével a Δ típus megtartása sikeres volt, a vad típusú U5 haplotípus szinte teljesen eliminálódott a sejtvonalból – a szekvenálási eredmények elemzése során az összes non-deléciós leolvasás között kb. 1% wt-U5 haplotípusút találtam. Így a továbbiakban minden, általam *wt*-ként megjelölt leolvasás egyben a wt-L3, HeLa sejtekből származó haplotípusra utal.

A változatok jelenlétét az aránylag rövid mitokondriális szekvenciákon könnyű vizuálisan ellenőrizni, ezért munkám során ezt rendszeresen megtettem. A heterozigótaként azonosított, ezáltal informatívnak ítélt változatok között a vizuális ellenőrzés során mindkét sejtvonal esetében találtam több olyat is, amely a várttól ellentétes haplotípus jellemzője volt, vagy haplotípuson belül mutatott eltérést (**16. ábra**). Például az egér NZB mitokondriális referenciához képest hívott két allél közül a nem-referencia allél volt jelen az NZB típusú readok 100%-ában. Az ilyen változatok a rekombinációt detektáló scriptem első változatában fals pozitívként jelentek meg. A vizuális ellenőrzési lépés lehetőséget adott a hiba felismerésére, majd a script következő verziójában mint figyelembe veendő „ellentétes” genotípus került beépítésére.



16. ábra – A deléció környéke a BH10 sejtvonalban (IGV). A szürke sáv mutatja a lefedettséget, a színes sávok variációkat jelölnek. Míg a jobb oldali három variáció valós különbségeket jelez a haplotípusok között – a piros, kék, illetve zöld színek arányosak a hívott bázisok mennyiségével az adott koordinátán – a bal oldali, a wt haplotípusban jelenlévő polimorfizmus a wt egységességéből adódóan egy tipikus hibásan hívott SNP.

Hasonló problémát jelentett először a nem egyértelmű változatok jelenléte. Ezek olyan SNP-k voltak, amelyek vagy mindkét típusú mtDNS szekvencián megjelentek néhány esetben, vagy csak az egyik haplotípusban, de nem minden ahhoz tartozó leolvasáson. Ezek kiszűrése kevésbé volt egyértelmű, mivel a két típus közül az első (mindkét haplotípusban előfordul) akár a rekombináció

jele is lehetne. Részben emiatt a kétértelműség miatt döntöttem később úgy, hogy amennyiben csak 1 SNP támogatta a rekombinációt, azt a leolvasást elvettem non-rekombinánsként, illetve a pontozási rendszerben negatív pontszámmal értékeltem. A második típusba tartozó SNP-eket (egy haplotípusban fordul elő, de csak az ahhoz tartozó leolvasások kis részében) teljes egészében elvettem.

A két egér haplotípus között eredetileg várt 90 SNP eltérésből 89-et valóban fel tudtam használni vizsgálataim során. A GATK HaplotypeCaller által ezek mind heterozigóta SNP-ként kerültek azonosításra, igazolva azon feltételezésem helyességét, hogy a heteroplazmikus sejtekben körülbelül azonos dózisban jelenlévő haplotípusokat diploidként hívva a különbségek heterozigótaként fognak megjelenni. Hasonló adat a két humán haplotípusról nem állt előzetesen rendelkezésemre. Azonos szűrési lépések után a BH10 sejtvonalon összesen 24 SNP-t és a hosszú deléciós szakaszt tudtam a szekvenciák haplotípus-hovatartozásának azonosítására felhasználni.

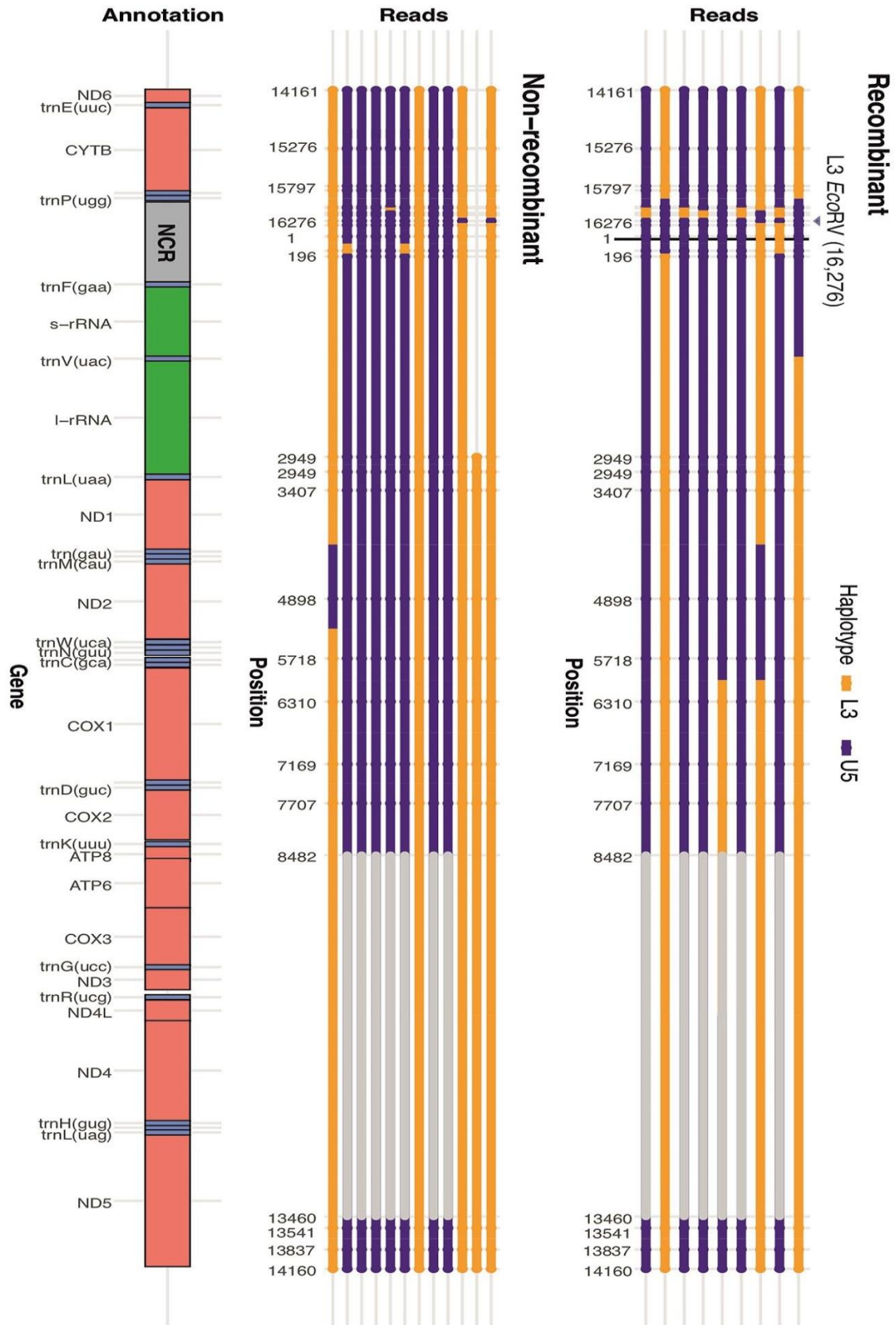
5.3.4. A rekombináció kimutatása mitokondriális DNS szekvenciákon

Az eredeti terv szerint a *wt* mtDNS haplotípusához tartozó SNP-k és a Δ mtDNS deléciójának egy szekvencián való jelenléte alapján mutattam volna ki rekombináció jelenlétét. Eredményeink alapján az, hogy mindegyik haplotípus esetében az SNP-kre támaszkodtam, több szempontból is jó döntésnek bizonyult. Egyrészt a rekombináció nem feltétlenül a deléciós szakasz környékén történt: a deléció hiánya, de a Δ típus SNP-inek jelenléte ugyanúgy jelenthet rekombinációt. Másrészt lehetetlen a nagy méretű deléció jelenlétére vagy hiányára építeni a HP202B sejtvonalon esetében, amely nem tartalmaz hasonlót. A deléció jelenléte végül a SNP-k típusával azonos súllyal lett számításba véve.

A pontozási rendszer beállításánál arra törekedtem, hogy lehetőség szerint ne kerüljenek fals pozitív eredmények a végső eredménybe, akkor sem, ha ezzel esetleg valós pozitív találatok kerülnek kizárásra. A 3-as minimum pontszám segít ezt megelőzni. Mivel a pontozási rendszer a lineáris leolvasásokon halad végig, és nem veszi figyelembe a cirkularitást, végeredményként a pozitív találatok legnagyobb eséllyel A-B-A vagy B-A-B mintázatot fognak adni, ahol A és B a két eltérő haplotípust jelzi. Ha mindhárom szakasz kellően hosszú, ez 3 pontot jelent. Annak, hogy a linearizált referencia végéhez olyan közel történik a rekombináció, hogy az végeredményben 2 pontot eredményez, van ugyan esélye, de mivel célom jelen esetben nem az összes rekombináns szekvencia megtalálása volt, hanem annak bizonyítása, hogy ezek jelen vannak, a nagyobb biztonságú találatok felé toltam az értékelést.

Ezzel együtt opcióként beépítettem a scriptbe a ponthatár és a pontozási séma kézi módosításának lehetőségét is. Alapbeállításaként jelenleg is a `--minscore 2` szerepel a scriptben: ez egy olyan pontszám, amely lineáris szekvenciák, vagy mtDNS esetében a teljes mitokondriális genom hosszánál rövidebb szekvenciák esetében már pozitív találatot jelentene, de az én célom szempontjából nem volt ideális. A BH10 sejtekben megfelelő pontszámot elért, illetve alatta maradó szekvenciákra mutat példákat a **17. ábra**.

A módszert párhuzamosan fejlesztettem és teszteltem a BH10 és HP202B sejtvonalon, ellenőrizve annak faj- és referenciafüggetlen alkalmasságát. A szükséges módosításokat, mint például a várttal „ellentétes” genotípusok figyelembevételét opcióként beépítettem. Nem megoldott viszont ezek kezelésének automatizálása: a detektáláshoz használt variációkat, ezek minőségét és alkalmazhatóságát egy új sejtvonalon esetében újra kell ellenőrizni. Mivel a végeredmény minősége nagyban függ a megbízható polimorfizmusok használatától, a variációk hívása és szűrése kiemelt fontosságú, ahogy a szekvenálás felhasználás előtti alapos minőségellenőrzése is.



17. ábra - Rekombinánsként (felül, Recombinant) és nem-rekombinánsként (középen, Non-recombinant) azonosított példa szekvenciák a BH10 sejtvonalból. Egy kék-sárga sáv egy leolvasásnak felel meg. Sárga és kék szín jelzi az egyes leolvasások haplotípusát. Az ábra alsó szekciója (Annotation) a mitokondriális genom annotációját mutatja. (Fragkoulis et al., 2024)

A leírt módszerrel és beállításokkal a BH10 sejtvonalon esetében 809, míg a HP202B sejtvonalon esetében 97 rekombináns leolvasást detektáltam összesen. A rekombinációs ráta ezzel 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).

A rekombináció teljes mennyisége mindkét sejtvonalon esetén konzervatív becslés. Nem csak választott elemzési paramétereimből adódóan, hanem abból is, hogy azonos haplotípus között előforduló rekombinációt sem laboratóriumi, sem bioinformatikai módszereinkkel nem tudunk detektálni.

5.3.5. A laboratóriumi eredmények sikeres igazolása bioinformatikai módszerrel

Munkám extenzív, több éves laboratóriumi munka megerősítésével szolgált, szekvenálási eredmények felhasználásával, ezen kívül egy olyan, leolvasás szintű rekombináció detektálási módszer, amelyre korábban nem állt rendelkezésre kidolgozott program. Módszerem konzisztensen előforduló polimorfizmusokra alapul. Nem teszi lehetővé a rekombináció nukleotid pontosságú helymeghatározását, de megmutatja, mely két SNP között történt rekombináció.

Bár a humán L3 és U5 mitokondriális haplotípusok között számos SNP eltérés adódik, ezek jelentős része a nem-kódoló régióban található. Az egymáshoz közel eső SNP-k, bár ebben a régióban részletes képet adnak, kevés információval szolgálnak az mtDNS többi szakaszáról. Emiatt is került bevonásra a HP202B sejtvonalon. Az NZB-Balb/c heteroplazmikus egereket korábban az emlős mitokondriális rekombináció megcáfolására is felhasználták (Hagstrom *et al.*, 2014). A rekombinációs ráta a mi kísérletünkben is igen alacsonynak adódott a HP202B sejtvonalonban. A BH10 sejtvonalonban ennek több, mint 10-szeresét figyeltük meg.

A rekombináció megfigyelésének gyakoriságában adódott jelentős különbségnek több oka is lehet. Ezek közül az egyik a HP202B sejtvonalon knockout részleges sikeressége. Ezen a nyilvánvaló különbségen kívül nem kizárható az sem, hogy a rekombinációt faj- és sejtvonalon-specifikus tényezők is befolyásolják (Dumont és Payseur, 2008). A HP202B sejtvonalon esetében azonban a többi hatást ellensúlyozhatja, hogy a rekombinációt bioinformatikailag könnyebb detektálni a szekvencián nagyjából egyenletesen elhelyezkedő polimorfizmusok segítségével.

5.3.6. A rekombináció jelenléte a mitokondriális genomban

A lineáris DNS-vezérelt mitokondriális rekombináció tehát lehetséges, bár megfigyelése még laboratóriumi körülmények között sem egyszerű. A mitokondriális genom fenntartása során a sejtek aránylag kevés energiát fordítanak javítási mechanizmusokra (Shokolenko *et al.*, 2013; Zhao, 2019), amely arra utal, hogy az nem „költséghatékony” a sejt szempontjából. Bár az állati sejteknek számos javítási mechanizmus rendelkezésére áll az mtDNS karbantartására (Fontana és Gahlon, 2020; Kazak *et al.*, 2012), a hatékony lineáris DNS degradációs mechanizmus arra utal, hogy a turnover a preferált megoldás, azaz a kettős száltörés javítása helyett a régi mtDNS helyettesítése újjal (Torregrosa-Muñumer *et al.*, 2019). Ezen kívül a kettős száltörés javítása potenciálisan hátrányos mutációkat eredményezhet (Fontana és Gahlon, 2020) akár NHEJ akár copy-choice rekombináció (Persson *et al.*, 2019; Phillips *et al.*, 2017) révén. Ugyanakkor feltételezhető, hogy a nagy mértékű oxidatív stressznek kitett szövetek esetében a sejtek több energiát fektetnek az mtDNS javításába (Herbers *et al.*, 2019; Pohjoismäki *et al.*, 2009).

A rekombináció egy univerzális mechanizmus az élővilág különböző területein. A DNS rekombináció része a gombák (Bágel'ová Poláková *et al.*, 2021; Fritsch *et al.*, 2014; Gerhold *et al.*, 2010; Ling, 2002) és növények (Barr *et al.*, 2005; Hecht *et al.*, 2011; Kozik *et al.*, 2019;

Maréchal és Brisson, 2010; Shedje *et al.*, 2007) mitokondriális biológiájának is. Ennek ellenére az mtDNS rekombinációjának lehetősége állatokban sokáig vitatott volt.

Az emlős mitokondriális DNS uniparentális, azaz egy szülőtől való öröklődése kevés lehetőséget ad új genetikai kombinációt eredményező rekombináció létrejöttére, arról nem is beszélve, hogy az szükségtelemé is teszi a mechanizmust. Mivel a legtöbb állati sejtben előforduló összes mitokondrium gyakorlatilag azonos génállománnyal rendelkezik, emiatt az esetlegesen előforduló rekombináció detektálása is problémás. Ezek a tényezők együttesen azt eredményezték, hogy sokáig általános feltételezés volt az állati sejtekben a mitokondriális rekombináció hiánya. Az első, annak jelenlétére utaló szekvenálási adatok esetében nehéz volt bizonyítani, hogy ezek valós eredmények – könnyen lehetnek volna a szekvenálási technológiából, a PCR amplifikációból és ligációból, vagy a rövid leolvasások téves összeszereléséből adódó artefaktok is (Ladoukakis és Zouros, 2017).

Korábbi kutatások során Kraytsberg és munkatársai (2004) humán heteroplazmiában 0,7% rekombináns DNS-t tudtak azonosítani, míg Zsurka és munkatársai (2005) 0,3% és 16% közé becsülték a rekombináns szekvenciák jelenlétét humán heteroplazmikus vázizomszövetekben. Saját, a HP202B sejtvonalon kapott 0,8%-os értékünk ehhez nagyságrendileg hasonló Kraytsberg és munkatársai (2004) által mért értéknek. A BH10 sejtvonalon detektált magasabb, 12%-os rekombinációs arányt az MGME1 teljesen sikeres kiütése mellett sikerült megfigyelni, míg a HP202B-ben továbbra is volt MGME1 expresszió – hasonlóan az említett, humán szövetekben és sejttenyészetekben végzett vizsgálatok esetéhez.

Az emlős mtDNS rekombinációjának lehetőségével szembeni szkepticizmus, illetve a lehetőség egyszerű elvetése más, mitokondriális témájú kutatásokra is hatással volt. Annak ellenére, hogy a cruciform (kereszt formájú) mtDNS intermedierek, vagy a Holliday-junction-ök jelenlétét baktériumok (Robinson, 2013) és élesztők (Zardoni *et al.*, 2020) nukleáris genomjának esetében a rekombináció jelének tekintették, ugyanezek létezése humán szívizomszövetben (Kajander *et al.*, 2001; Pohjoismäki *et al.*, 2009), eger szövetekben (Pohjoismäki *et al.*, 2013), vagy épp sejt kultúrákban (Torregrosa-Muñumer *et al.*, 2015) nem váltott ki különösebb tudományos érdeklődést azok szerepe iránt.

Saját kutatásaink is bizonyították, hogy bár nem gyakori, vadon élő állatokban is előfordulnak heteroplazmikus egyedek. Tapanainen *et al.* (2024) munkája során finnországi *Lepus* fajok mitokondriumának long repeat (LR) szakaszát vizsgálva 3%-ban talált heteroplazmikus egyedeket a havasi nyúl esetében, és 12%-ban a mezei nyúléban (151 és 148 minta). A két faj hibrid zónájában ezen kívül számos esetben fordult elő mitokondriális introgresszió is a két faj között (Pohjoismäki *et al.*, 2021). Így, annak ellenére, hogy a sikeres, biológiai szempontból jelentős mitokondriális rekombináció esélye igen alacsony, nem zárható ki annak megtörténte és esetleges befolyása a fajok környezethez való alkalmazkodására sem.

6. KÖVETKEZTETÉSEK ÉS JAVASLATOK

6.1. Domesztikáció és szelekció üregi nyúlban

6.1.1. Házi- és vad üregi nyúl állományok hibridizációja

A hibridizáció az üregi nyúl vad és házasított formái között több esetben is előfordul. A franciaországi minták esetében a poolozott mintákat vizsgáltunk, így a minták vegyes genetikai hátterénél többet nem tudunk megállapítani, a magyarországi esetekben viszont egyedenként ki tudtuk mutatni a hibridizációt (5. *ábra*).

A magyarországi üregi nyúl az ibériai és francia ágaktól genetikailag jól elkülöníthető populáció (2. *ábra* és 4. *ábra*). A házinyúllal hibridnek adódott egyedek közül mind a Budapesti Állat- és Növénykertben tartott és mintázott populációból származott. Ezekben az egyedekben a ROH szakaszok hossza és mennyisége is magasabb volt (7. *ábra*). A Tiszaszentimre környékén gyűjtött minták nem adódtak hibridnek, és a genomok ROH szakaszai sem adódtak kiemelkedőnek, sem mennyiségük, sem hosszuk szempontjából. Ugyanakkor nem áll rendelkezésünkre információ más, Magyarországon vadon élő populációk állapotáról – könnyen előfordulhat hibridizáció, vagy súlyosabb beltenyésztettség is, amelyet érdemes lenne a jövőben részletesebben feltérképezni.

6.1.2. Poligenikus domesztikáció és szabályozó szakaszok szerepe

A domesztikációs folyamat számos fajban végzett számos vizsgálatok alapján sokgénés (Carneiro *et al.*, 2014b; Ramos-Onsins *et al.*, 2014; Rochus *et al.*, 2018; Zheng *et al.*, 2020, stb.), elsősorban az allélgyakoriságok eltolódásán alapul, amellyel munkám eredményei is megerősítik. Eredményeim hasonlóan adódtak a szelektált gének funkcióiban is, bár kevés esetben adódtak szelektáltnak pontosan ugyanazok a gének.

A genom nem-kódoló szakaszaiban kevésbé vizsgált a szelekció. Ezen régiók funkciója, szerepe is kevésbé ismert. Expressziós vizsgálatok alapján a géneken belüli mutációkon kívül az expressziós szint megváltozása is fontos szerepet játszhat a domesztikációban (Sato *et al.*, 2020). Saját eredményeim is erre utalnak, a szelektált szakaszokban adódott egyes TFBS típusok dúsulása alapján (4. *táblázat*).

Célszerűnek tartom a nem-kódoló, intronikus, UTR és intergénikus szakaszok alaposabb megfontolását szelekciós folyamatok vizsgálatában, azok jobbra ismeretlen, de feltehetően jelentős szerepe miatt. A szelekciós vizsgálatok megismétlése, illetve kibővítése nagyobb folytonosságú, megbízhatóbb, és jobban annotálható referenciagenom felhasználásával együtt szintén elősegítheti e folyamatok megértését.

6.2. A havasi nyúl (*Lepus timidus*) referenciagenom

A havasi nyúl fontos ökológiai szerepet tölt be élőhelyének táplálékhálózatában, emellett ikonikus faj, amely a hideg, havas éghajlathoz alkalmazkodott, és ennek megfelelő egyedi fiziológiai és morfológiai jellemzőkkel rendelkezik. A faj genetikáját, genomikáját és evolúcióját számos tanulmány vizsgálta eddig is (Melo-Ferreira *et al.*, 2009; Smith *et al.*, 2017; Levänen *et al.*, 2018a; 2018b; Ferreira *et al.*, 2021; Pohjoismäki *et al.*, 2021a; Giska *et al.*, 2022; Michell *et al.*, 2022; Gaertner *et al.*, 2023), az általam összeszerelt genom pedig további vizsgálatok alapjául szolgálhat.

A jó minőségű, génannotációval rendelkező referenciagenomok lehetővé teszik a reverz genetikai kísérleteket, egyes gének közvetlen manipulációját és annak hatásának megfigyelését. Így lehetséges egyes változatok tulajdonságokra gyakorolt hatásának vizsgálatát, amelyekre máskülönben populációgenetikai adatokból kéne következtetni.

Az újonnan összeszerelt referenciák megkönnyítik a finnországi havasi és mezei nyulak összehasonlító vizsgálatát is. Rövid leolvasásos szekvenálási adatok mindkét genomra illesztése során mindegyikre a leolvasások 100%-át sikerült illeszteni. Ezek ddRAD szekvenálási technológiával nyert adatok, amelyek a genomot csak részleteiben reprezentálják, és a fajok közötti különbségek egyes restriktációs pontok elvesztéséhez vezethetnek (Arnold *et al.*, 2013; Davey *et al.*, 2013). Ugyanakkor így is jó képet kaphatunk a genomok felhasználhatóságáról a két faj összehasonlításában.

Hasonló megbízhatóságú eredményeket kaptunk mindkét fajból. A variációk mennyiségét tekintve jellemzően a referenciaazonos fajban számszerűen kevesebb az eltérés, és ezeknek arányaiban kisebb része inszerció vagy delécio (saját nem publikált adat).

A finnországi és ír havasi nyúl populációk genetikai különbözősége korábban is ismert volt, de a pontosabb referenciagenom felhasználható részletesebb, populáció szintű genomi összehasonlításokhoz, és az alfajok közötti szelekció és a fajképzés mechanizmusának vizsgálatához. Veszélyeztetett alfajok, mint a *Lepus timidus sylvaticus* (Thulin *et al.*, 2021; Michell *et al.*, 2022), vagy a *Lepus timidus hibernicus* (Reid, 2011; Caravaggi *et al.*, 2017) konzervációjában is szerepet tölthet be a fontos variációk és a köztük lévő kapcsoltság felderítése.

6.3. Mitokondriális rekombináció alátámasztása emlős sejtvonalakban

Az emlős sejtekben előforduló mitokondriális rekombináció lehetőségét sikeresen alátámasztottam egy új, nagy pontosságú, hosszú leolvasásos szekvenáláson alapuló bioinformatikai módszerrel. Módszerem molekuláris biológiai kísérletek és transzmissziós elektronmikroszkópiás megfigyelések eredményeit tudta megerősíteni (Fragkoulis *et al.*, 2024).

Bár a vizsgálatához használt sejtvonalak specifikusan a célhoz választottak voltak, hogy a detektálás esélyét megnöveljük, az emlős mtDNS rekombinációjának lehetősége nem kizárt természetes körülmények között. Az mtDNS maternális öröklődése és túlnyomó részt homoplazmikus előfordulása miatt annak esélye alacsony, ugyanakkor nem kizárható. Egymással hibridizálódó *Lepus* fajokban például sűrűn megfigyelt a heteroplazmia és a mitokondriális introgresszió is (Levänen *et al.*, 2018a; Tapanainen *et al.*, 2024). Biparentális öröklődésű mtDNS esetében pedig már ismert evolúciós időskálán bekövetkezett rekombináció bekövetkezése (Ladoukakis and Zouros, 2001; Ladoukakis *et al.*, 2011).

A fentiek alapján javaslom az mtDNS haplotípusok alapján történő vizsgálatokkor az introgresszió, és a rekombináció lehetőségének megfontolását, különösen, ha az magyarázattal szolgálhat egyes anomáliákra, főleg, ha ismert a populációban a heteroplazmia, vagy a gyakori mtDNS introgresszió előfordulása.

7. 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 háziasított 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 génekkel 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.

8. ÖSSZEFOGLALÁS

8.1. Bevezetés

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* nemzetségbe számos ma élő fajának képviselői világszerte megtalálhatóak. A nemzetség széles körű elterjedésének, és bizonyos esetekben emberrel való asszociációjuk (Somerville *et al.*, 2017; Sheng *et al.*, 2020) 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 (Nevado *et al.*, 2014; Bohling, 2020). 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, amelyekben heteroplazmikusság is aránylag gyakori – például *Lepus timidus* és *Lepus europaeus* (Tapanainen *et al.*, 2024) esetében – nem kizárható az esetleges, evolúciós szempontból jelentős rekombináció sem.

8.2. Anyag és módszer

8.2.1. 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 beltenyésztettség 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 (Fst) 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 transzkripció faktor kötőhelyek (TFBS) szempontjából vizsgáltam. A gének azonosítását az Ensembl (release 104) (Howe, K.L. *et al.*, 2021) annotációja alapján végeztem, TFBS-ek meghatározását pedig LiftOverrel (Hinrichs, 2006), a ChipSummitDB (Czipa *et al.*, 2020) adatbázis alapján.

8.2.2. *Lepus timidus* referenciagenom összeszerelése

A genomhoz 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 fülszövetből. A mtDNS PCR amplifikációt követően Illumina módszerrel volt szekvenálva.

A genom összeszerelésekor a modern sztenderdeket (Rhie *et al.*, 2020a) 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 tiddk-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 QUAST (Gurevich *et al.*, 2013) – szekvencia tulajdonságok és statisztikák – programokat használtam.

8.2.3. 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 egér (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 rendszerállítottam fel. Valamennyi lépést párhuzamosan végeztem a két sejtvonalon, vizuális ellenőrzésekkel együtt. A célra írt scriptek elérhetők a <https://zenodo.org/records/10473729> címen.

8.3. Eredmények és diszkusszió

8.3.1. 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 (**4. 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.

A házi- és üregi nyúl hibridizációja több magyar üregi nyúl mintában is megfigyelhető volt, ezek mindegyike egyazon szubpopulációhoz tartozik. A hibridizáció konzisztensnek adódott a különböző tesztek során, és a hibrid állatok mindegyike az állatkerti populációba tartozott (**5. ábra** és **6. ábra**, **M6. Melléklet**). Ugyanakkor nem áll rendelkezésünkre információ arról, hogy honnan, hogyan és mikor történhetett pontosan a hibridizáció. 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.*, 2014b), allélfrekvencia eltolódásokon alapul, nem pedig új mutációk megjelenésén és elterjedésén. Az üregi- és házinyúl közötti differenciációt és szelekciót három módszerrel megvizsgálva a szelektálnak adódott szakaszokkal (összesen 46 szakasz) átfedő gének (**2. táblázat**) között egyedfejlődéssel, növekedéssel, jelátvitellel – viselkedéssel – és immunválasszal kapcsolatosak is adódtak. Ezek olyan tulajdonságokat befolyásolnak, amelyek jellemzően megváltoznak a domesztikáció során.

*A kiválasztott régiókban nagy mennyiségű erősen differenciált ($F_{st} > 0.75$, 983 variáció, **5. táblázat**), de kevés teljesen fixált változat adódott. Emellett adódtak olyan szakaszok is (14), amelyekben nincs gén annotálva (*

3. táblázat). Megvizsgáltuk a transzkripció faktor kötőhelyek (TFBS) jelenlétét is, és egyes típusok dúsulását tapasztaltuk a szelektált szakaszokban (**4. táblázat**), a szabályozó régiók fontosságát mutatva.

8.3.2. 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 becsültnél (23%) (Marques *et al.*, 2020). Ez valószínűleg annak köszönhető, hogy a hosszú leolvasásos szekvenálás 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 (**12. 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 (**14. 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 (**M7. Melléklet**). Az összeszerelt mitokondriális szekvencia 17 482 bázis hosszúságúnak adódott, ezen felül cirkuláris, és minden várt gén teljes egészében annotált rajta (**13. ábra**).

8.3.3. 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 (**15. ábra**).

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 illesztését követően BH10 sejtvonalba 24 SNP-t és a hosszú deléciós szakaszt, a HP202B sejtvonalba pedig 89 SNP-t azonosítottam, amely megbízhatóan differenciálnak 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 (**M5. melléklet**) és pontozási rendszerrel végeztem. Az alkalmazott módszerrel és beállításokkal a rekombinációs leolvasások aránya a BH10 sejtvonalba 12%-nak (809 rekombináns / 6730 tisztán mitokondriális eredetű leolvasás), a HP202B sejtvonalba pedig 0,8%-nak adódott (97 rekombináns / 11976 tisztán mitokondriális eredetű leolvasás).

8.4. Következtetések és javaslatok

- Bár a tizzaszentimrei, vadon élő populáció nem adódott házinyúllal hibridizáltak, 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 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 lehetőségessége 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.

9. SUMMARY

9.1. Introduction

Leporidae species can be found all over the world, populating diverse 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 domesticated and wild forms. On the other hand, the genus *Lepus* consists of numerous extant species worldwide. Despite the overall 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 are able to 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 uniparentally 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.

9.2. Materials and methods

9.2.1. 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, totally 100 samples. Our wild rabbit samples are from Hungary, France, and the Iberian Peninsula, the first group sequenced by us, while the latter two were collected from the NCBI SRA 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 populations, using the ADMIXTURE (Alexander *et al.*, 2009) and Treemix (Pickrell and Pritchard, 2012) programs. The 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 (99th 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) (Howe, K. L. *et al.*, 2021), while transcription binding sites were identified using LiftOver (Hinrichs, 2006) and the ChipSummitDB (Czipa *et al.*, 2020) database.

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

The mountain hare 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. Nuclear DNA for the PacBio HiFi and Illumina HiC genome sequencing was isolated from a 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 masked using RepeatModeler (Flynn *et al.*, 2020) and RepeatMasker (Smit *et al.*, 2015), while telomer sequences were identified with tiddk (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.

9.2.3. 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 (A or B) I generated a pseudo-sequence and set up a 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>.

9.3. Results and discussion

9.3.1. 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 4.**), suggesting their natural origins rather than being descendants of released or escaped domestic animals.

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 (**Figures 5. and 6.** and [Supplement M4](#)), 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 (**Table 2.**) 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 ($F_{st} > 0.75$) variants (983 in total, detailed in **Table 5.**) 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, **Table 3.**) 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 4.**), showing the importance of regulatory regions.

9.3.2. 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 12.**). 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). 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 14.**), 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) ([Supplementary M5](#)). The assembled mitochondrial DNA sequence was 17 482 bases, circular, and we were able to annotate every expected gene on it (**Figure 13.**).

9.3.3. 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 15.**).

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 ([Supplementary M3](#)). 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.

9.4. Conclusions and suggestions

- Even though the free-living Hungarian rabbit population (Tizzaszentimre) 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 possibly explain by it.

MELLÉKLETEK

M1. IRODALOMJEGYZÉK

- Adessi, C. (2000). Solid phase DNA amplification: Characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Research*, 28(20), 87e–887. <https://doi.org/10.1093/nar/28.20.e87>
- 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>
- Alves, J. M., Carneiro, M., Afonso, S., Lopes, S., Garreau, H., Boucher, S., Allain, D., Queney, G., Esteves, P. J., Bolet, G., & Ferrand, N. (2015). Levels and Patterns of Genetic Diversity and Population Structure in Domestic Rabbits. *PloS One*, 10(12), e0144687. <https://doi.org/10.1371/journal.pone.0144687>
- Alves, J. M., Carneiro, M., Day, J. P., Welch, J. J., Duckworth, J. A., Cox, T. E., Letnic, M., Strive, T., Ferrand, N., & Jiggins, F. M. (2022). A single introduction of wild rabbits triggered the biological invasion of Australia. *Proceedings of the National Academy of Sciences of the United States of America*, 119(35), e2122734119. <https://doi.org/10.1073/pnas.2122734119>
- Alves, P. C., Ferrand, N., Suchentrunk, F., & Harris, D. J. (2003). Ancient introgression of *Lepus timidus* mtDNA into *L. granatensis* and *L. europaeus* in the Iberian Peninsula. *Molecular Phylogenetics and Evolution*, 27(1), 70–80. [https://doi.org/10.1016/S1055-7903\(02\)00417-7](https://doi.org/10.1016/S1055-7903(02)00417-7)
- Alves, P. C., Melo-Ferreira, J., Freitas, H., & Boursot, P. (2008). The ubiquitous mountain hare mitochondria: Multiple introgressive hybridization in hares, genus *Lepus*. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1505), 2831–2839. <https://doi.org/10.1098/rstb.2008.0053>
- Anderson, C. M., Chen, S. Y., Dimon, M. T., Oke, A., DeRisi, J. L., & Fung, J. C. (2011). ReCombine: A suite of programs for detection and analysis of meiotic recombination in whole-genome datasets. *PloS One*, 6(10), e25509. <https://doi.org/10.1371/journal.pone.0025509>
- Andersson, L., & Purugganan, M. (2022). Molecular genetic variation of animals and plants under domestication. *Proceedings of the National Academy of Sciences of the United States of America*, 119(30), e2122150119. <https://doi.org/10.1073/pnas.2122150119>
- Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. [Computer software]. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Angerbjorn, A., & Fluxus, J. E. C. (1995). *Lepus timidus*. *Mammalian Species*, 495, 1. <https://doi.org/10.2307/3504302>

- Arnold, B., Corbett-Detig, R. B., Hartl, D., & Bomblies, K. (2013). RAD seq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, 22(11), 3179–3190. <https://doi.org/10.1111/mec.12276>
- 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.
- Bacman, S. R., Williams, S. L., Garcia, S., & Moraes, C. T. (2010). Organ-specific shifts in mtDNA heteroplasmy following systemic delivery of a mitochondria-targeted restriction endonuclease. *Gene Therapy*, 17(6), 713–720. <https://doi.org/10.1038/gt.2010.25>
- Bágel'ová Poláková, S., Lichtner, Ž., Szemes, T., Smolejová, M., & Sulo, P. (2021). Mitochondrial DNA duplication, recombination, and introgression during interspecific hybridization. *Scientific Reports*, 11(1), 12726. <https://doi.org/10.1038/s41598-021-92125-y>
- Bai, F., Cai, Y., Qi, M., Liang, C., Pan, L., Liu, Y., Feng, Y., Cao, X., Yang, Q., Ren, G., Jiao, S., Gao, S., Lu, M., Wang, X., Heller, R., Lenstra, J. A., & Jiang, Y. (2025). *LCORL* and *STC2* Variants Increase Body Size and Growth Rate in Cattle and Other Animals. *Genomics, Proteomics & Bioinformatics*, qzaf025. <https://doi.org/10.1093/gpbjnl/qzaf025>
- Balboa, D., Weltner, J., Euroola, S., Trokovic, R., Wartiovaara, K., & Otonkoski, T. (2015). Conditionally Stabilized dCas9 Activator for Controlling Gene Expression in Human Cell Reprogramming and Differentiation. *Stem Cell Reports*, 5(3), 448–459. <https://doi.org/10.1016/j.stemcr.2015.08.001>
- Ballan, M., Bovo, S., Bertolini, F., Schiavo, G., Schiavitto, M., Negrini, R., & Fontanesi, L. (2023). Population genomic structures and signatures of selection define the genetic uniqueness of several fancy and meat rabbit breeds. *Journal of Animal Breeding and Genetics = Zeitschrift Fur Tierzucht Und Zuchtungsbiologie*, 140(6), 663–678. <https://doi.org/10.1111/jbg.12818>
- Ballan, M., Bovo, S., Schiavo, G., Schiavitto, M., Negrini, R., & Fontanesi, L. (2022). Genomic diversity and signatures of selection in meat and fancy rabbit breeds based on high-density marker data. *Genetics, Selection, Evolution: GSE*, 54(1), 3. <https://doi.org/10.1186/s12711-022-00696-9>
- Barr, C. M., Neiman, M., & Taylor, D. R. (2005). Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist*, 168(1), 39–50. <https://doi.org/10.1111/j.1469-8137.2005.01492.x>
- Barton, N. H. (1998). The effect of hitch-hiking on neutral genealogies. *Genetical Research*, 72(2), 123–133. <https://doi.org/10.1017/S0016672398003462>
- Battersby, B. J. (2001). Selection of a mtDNA sequence variant in hepatocytes of heteroplasmic mice is not due to differences in respiratory chain function or efficiency of replication. *Human Molecular Genetics*, 10(22), 2469–2479. <https://doi.org/10.1093/hmg/10.22.2469>
- Bay, R. A., & Bielawski, J. P. (2011). Recombination Detection Under Evolutionary Scenarios Relevant to Functional Divergence. *Journal of Molecular Evolution*, 73(5–6), 273–286. <https://doi.org/10.1007/s00239-011-9473-0>
- Bayona-Bafaluy, M. P., Blits, B., Battersby, B. J., Shoubridge, E. A., & Moraes, C. T. (2005). Rapid directional shift of mitochondrial DNA heteroplasmy in animal tissues by a mitochondrially targeted restriction endonuclease. *Proceedings of the National Academy of Sciences*, 102(40), 14392–14397. <https://doi.org/10.1073/pnas.0502896102>
- 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. M., 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 Research*, *19*(4), 549–565. <https://doi.org/10.1007/s10577-011-9210-y>
- 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. <https://doi.org/10.1016/j.ymeth.2017.04.004>
- Bello-Rodríguez, V., Mateo, R. G., Pellissier, L., Cubas, J., Cooke, B., & González-Mancebo, J. M. (2021). Forecast increase in invasive rabbit spread into ecosystems of an oceanic island (Tenerife) under climate change. *Ecological Applications: A Publication of the Ecological Society of America*, *31*(1), e02206. <https://doi.org/10.1002/eap.2206>
- Berg, J. J., & Coop, G. (2014). A Population Genetic Signal of Polygenic Adaptation. *PLoS Genetics*, *10*(8), e1004412. <https://doi.org/10.1371/journal.pgen.1004412>
- Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A., Aury, J.-M., Louis, A., Dehais, P., Bardou, P., Montfort, J., Klopp, C., Cabau, C., Gaspin, C., Thorgaard, G. H., ... Guiguen, Y. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications*, *5*(1), 3657. <https://doi.org/10.1038/ncomms4657>
- Besnier, F., Ayllon, F., Skaala, Ø., Solberg, M. F., Fjeldheim, P. T., Anderson, K., Knutar, S., & Glover, K. A. (2022). Introgression of domesticated salmon changes life history and phenology of a wild salmon population. *Evolutionary Applications*, *15*(5), 853–864. <https://doi.org/10.1111/eva.13375>
- Biscotti, M. A., Olmo, E., & Heslop-Harrison, J. S. (2015). Repetitive DNA in eukaryotic genomes. *Chromosome Research*, *23*(3), 415–420. <https://doi.org/10.1007/s10577-015-9499-z>
- Bock, A. (2020). *Lepus europaeus* (Lagomorpha: Leporidae). *Mammalian Species*, *52*(997), 125–142. <https://doi.org/10.1093/mspecies/seaa010>
- 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>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bősze, Z., Major, P., Baczkó, I., Odening, K. E., Bodrogi, L., Hiripi, L., & Varró, A. (2016). The potential impact of new generation transgenic methods on creating rabbit models of cardiac diseases. *Progress in Biophysics and Molecular Biology*, *121*(2), 123–130. <https://doi.org/10.1016/j.pbiomolbio.2016.05.007>
- Braslavsky, I., Hebert, B., Kartalov, E., & Quake, S. R. (2003). Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciences*, *100*(7), 3960–3964. <https://doi.org/10.1073/pnas.0230489100>
- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S. R., Moon, K., Burcham, T., Pallas, M., ... Corcoran, K. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology*, *18*(6), 630–634. <https://doi.org/10.1038/76469>

- 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>
- Browning, S. R., Waples, R. K., & Browning, B. L. (2023). Fast, accurate local ancestry inference with FLARE. *The American Journal of Human Genetics*, *110*(2), 326–335. <https://doi.org/10.1016/j.ajhg.2022.12.010>
- Brusini, I., Carneiro, M., Wang, C., Rubin, C.-J., Ring, H., Afonso, S., Blanco-Aguiar, J. A., Ferrand, N., Rafati, N., Villafuerte, R., Smedby, Ö., Damberg, P., Hallböök, F., Fredrikson, M., & Andersson, L. (2018). Changes in brain architecture are consistent with altered fear processing in domestic rabbits. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(28), 7380–7385. <https://doi.org/10.1073/pnas.1801024115>
- Butlin, R. K. (2005). Recombination and speciation. *Molecular Ecology*, *14*(9), 2621–2635. <https://doi.org/10.1111/j.1365-294X.2005.02617.x>
- Carneiro, M., Afonso, S., Geraldes, A., Garreau, H., Bolet, G., Boucher, S., Tircazes, A., Queney, G., Nachman, M. W., & Ferrand, N. (2011). The genetic structure of domestic rabbits. *Molecular Biology and Evolution*, *28*(6), 1801–1816. <https://doi.org/10.1093/molbev/msr003>
- Carneiro, M., Albert, F. W., Afonso, S., Pereira, R. J., Burbano, H., Campos, R., Melo-Ferreira, J., Blanco-Aguiar, J. A., Villafuerte, R., Nachman, M. W., Good, J. M., & Ferrand, N. (2014a). The genomic architecture of population divergence between subspecies of the European rabbit. *PLoS Genetics*, *10*(8), e1003519. <https://doi.org/10.1371/journal.pgen.1003519>
- Carneiro, M., Hu, D., Archer, J., Feng, C., Afonso, S., Chen, C., Blanco-Aguiar, J. A., Garreau, H., Boucher, S., Ferreira, P. G., Ferrand, N., Rubin, C.-J., & Andersson, L. (2017). Dwarfism and Altered Craniofacial Development in Rabbits Is Caused by a 12.1 kb Deletion at the HMGA2 Locus. *Genetics*, *205*(2), 955–965. <https://doi.org/10.1534/genetics.116.196667>
- Carneiro, M., Piorno, V., Rubin, C.-J., Alves, J. M., Ferrand, N., Alves, P. C., & Andersson, L. (2015). Candidate genes underlying heritable differences in reproductive seasonality between wild and domestic rabbits. *Animal Genetics*, *46*(4), 418–425. <https://doi.org/10.1111/age.12299>
- Carneiro, M., Rubin, C.-J., Di Palma, F., Albert, F. W., Alföldi, J., Barrio, A. M., 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. (2014b). Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science*, *345*(6200), 1074–1079. <https://doi.org/10.1126/science.1253714>
- Carvajal-Rodríguez, A. (2017). HacDivSel: Two new methods (haplotype-based and outlier-based) for the detection of divergent selection in pairs of populations. *PloS One*, *12*(4), e0175944. <https://doi.org/10.1371/journal.pone.0175944>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Chaisson, M. J. P., Huddleston, J., Dennis, M. Y., Sudmant, P. H., Malig, M., Hormozdiari, F., Antonacci, F., Surti, U., Sandstrom, R., Boitano, M., Landolin, J. M., Stamatoyannopoulos, J. A., Hunkapiller, M. W., Korlach, J., & Eichler, E. E. (2015). Resolving the complexity

- of the human genome using single-molecule sequencing. *Nature*, *517*(7536), 608–611. <https://doi.org/10.1038/nature13907>
- Challis, R., Richards, E., Rajan, J., Cochrane, G., & Blaxter, M. (2020). BlobToolKit – Interactive Quality Assessment of Genome Assemblies. *G3 Genes|Genomes|Genetics*, *10*(4), 1361–1374. <https://doi.org/10.1534/g3.119.400908>
- 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>
- Chen, Z. (2022). Robust tests for combining p-values under arbitrary dependency structures. *Scientific Reports*, *12*(1), 3158. <https://doi.org/10.1038/s41598-022-07094-7>
- Cheng, H., Asri, M., Lucas, J., Koren, S., & Li, H. (2024). Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. *Nature Methods*, *21*(6), 967–970. <https://doi.org/10.1038/s41592-024-02269-8>
- 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>
- Chikhi, R., & Medvedev, P. (2014). Informed and automated k -mer size selection for genome assembly. *Bioinformatics*, *30*(1), 31–37. <https://doi.org/10.1093/bioinformatics/btt310>
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92. <https://doi.org/10.4161/fly.19695>
- Codner, G. F., Erbs, V., Loeffler, J., Chessum, L., Caulder, A., Jullien, N., Wells, S., Birling, M.-C., & Teboul, L. (2021). Universal Southern blot protocol with cold or radioactive probes for the validation of alleles obtained by homologous recombination. *Methods*, *191*, 59–67. <https://doi.org/10.1016/j.ymeth.2020.06.011>
- Crisci, J. L., Poh, Y.-P., Mahajan, S., & Jensen, J. D. (2013). The impact of equilibrium assumptions on tests of selection. *Frontiers in Genetics*, *4*. <https://doi.org/10.3389/fgene.2013.00235>
- Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). THE HIDDEN SIDE OF INVASIONS: MASSIVE INTROGRESSION BY LOCAL GENES. *Evolution*, ???-???. <https://doi.org/10.1111/j.1558-5646.2008.00413.x>
- 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>
- Dai, M., Sun, H., Zhao, L., Wu, Q., You, B., Xu, F., Liao, J., Zhu, S., Li, Z., Yao, Y., Nair, V., & Liao, M. (2022). Duck CD8⁺ T Cell Response to H5N1 Highly Pathogenic Avian Influenza Virus Infection In Vivo and In Vitro. *The Journal of Immunology*, *209*(5), 979–990. <https://doi.org/10.4049/jimmunol.2101147>
- Dai, Y., Hu, S., Bai, S., Li, J., Yang, N., Zhai, P., Zhao, B., Chen, Y., & Wu, X. (2022). CDK1 promotes the proliferation of melanocytes in Rex rabbits. *Genes & Genomics*, *44*(10), 1191–1199. <https://doi.org/10.1007/s13258-022-01283-4>

- 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>
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. <https://doi.org/10.1093/gigascience/giab008>
- Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., & Blaxter, M. L. (2013). Special features of RAD Sequencing data: Implications for genotyping. *Molecular Ecology*, 22(11), 3151–3164. <https://doi.org/10.1111/mec.12084>
- Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology*, 34(5), 518–524. <https://doi.org/10.1038/nbt.3423>
- DeAngelis, M. M., Wang, D. G., & Hawkins, T. L. (1995). Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Research*, 23(22), 4742–4743. <https://doi.org/10.1093/nar/23.22.4742>
- 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>
- Delibes, R., & Delibes-Mateos, M. (2015). Linking historical ecology and invasion biology: Some lessons from European rabbit introductions into the new world before the nineteenth century. *Biological Invasions*, 17(9), 2505–2515. <https://doi.org/10.1007/s10530-015-0905-4>
- Demars, J., Iannuccelli, N., Utzeri, V. J., Auvinet, G., Riquet, J., Fontanesi, L., & Allain, D. (2018). New Insights into the Melanophilin (MLPH) Gene Affecting Coat Color Dilution in Rabbits. *Genes*, 9(9), 430. <https://doi.org/10.3390/genes9090430>
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491–498. <https://doi.org/10.1038/ng.806>
- Dumont, B.L, Payseur, B. A., (2008) EVOLUTION OF THE GENOMIC RATE OF RECOMBINATION IN MAMMALS, *Evolution*, Volume 62, Issue 2, 1 February 2008, Pages 276–294, <https://doi.org/10.1111/j.1558-5646.2007.00278.x>
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for Ancient Admixture between Closely Related Populations. *Molecular Biology and Evolution*, 28(8), 2239–2252. <https://doi.org/10.1093/molbev/msr048>
- Elbakry, A., & Löbrich, M. (2021). Homologous Recombination Subpathways: A Tangle to Resolve. *Frontiers in Genetics*, 12, 723847. <https://doi.org/10.3389/fgene.2021.723847>
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
- Fan, J., Chen, Y., Yan, H., Liu, B., Wang, Y., Zhang, J., Chen, Y. E., Liu, E., & Liang, J. (2018). Genomic and Transcriptomic Analysis of Hypercholesterolemic Rabbits: Progress and Perspectives. *International Journal of Molecular Sciences*, 19(11), 3512. <https://doi.org/10.3390/ijms19113512>

- Fan, S., Hansen, M. E. B., Lo, Y., & Tishkoff, S. A. (2016). Going global by adapting local: A review of recent human adaptation. *Science*, 354(6308), 54–59. <https://doi.org/10.1126/science.aaf5098>
- Fatima, N., Jia, L., Liu, B., Li, L., Bai, L., Wang, W., Zhao, S., Wang, R., & Liu, E. (2023). A homozygous missense mutation in the fibroblast growth factor 5 gene is associated with the long-hair trait in Angora rabbits. *BMC Genomics*, 24(1), 298. <https://doi.org/10.1186/s12864-023-09405-2>
- Fekete, Z., Absolon, D. E., Michell, C., Wood, J. M. D., Goffart, S., & Pohjoismäki, J. L. O. (2025a). Chromosome-level reference genome assembly for the mountain hare (*Lepus timidus*). *Peer Community Journal*, 5, e14. <https://doi.org/10.24072/pcjournal.514>
- Fekete, Z., Németh, Z., Ninausz, N., Fehér, P., Schiller, M., Alnajjar, M., Szenes, Á., Nagy, T., Stéger, V., Kontra, L., & Barta, E. (2025b). Whole-Genome Sequencing-Based Population Genetic Analysis of Wild and Domestic Rabbit Breeds. *Animals*, 15(6), 775. <https://doi.org/10.3390/ani15060775>
- Ferreira, M. S., Jones, M. R., Callahan, C. M., Farelo, L., Tolesa, Z., Suchentrunk, F., Boursot, P., Mills, L. S., Alves, P. C., Good, J. M., & Melo-Ferreira, J. (2021). The Legacy of Recurrent Introgression during the Radiation of Hares. *Systematic Biology*, 70(3), 593–607. <https://doi.org/10.1093/sysbio/syaa088>
- Ferrer-Admetlla, A., Liang, M., Korneliussen, T., & Nielsen, R. (2014). On Detecting Incomplete Soft or Hard Selective Sweeps Using Haplotype Structure. *Molecular Biology and Evolution*, 31(5), 1275–1291. <https://doi.org/10.1093/molbev/msu077>
- 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>
- Fontana, G. A., & Gahlon, H. L. (2020). Mechanisms of replication and repair in mitochondrial DNA deletion formation. *Nucleic Acids Research*, 48(20), 11244–11258. <https://doi.org/10.1093/nar/gkaa804>
- 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>
- Fragkoulis, G., Hangan, A., Fekete, Z., Michell, C., Moraes, C. T., Willcox, S., Griffith, J. D., Goffart, S., & 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>
- Fritsch, E. S., Chabbert, C. D., Klaus, B., & Steinmetz, L. M. (2014). A Genome-Wide Map of Mitochondrial DNA Recombination in Yeast. *Genetics*, 198(2), 755–771. <https://doi.org/10.1534/genetics.114.166637>
- Gaertner, K., Michell, C., Tapanainen, R., Goffart, S., Saari, S., Soininmäki, M., Dufour, E., & Pohjoismäki, J. L. O. (2023). Molecular phenotyping uncovers differences in basic

- housekeeping functions among closely related species of hares (*Lepus* spp., Lagomorpha: Leporidae). *Molecular Ecology*, 32(15), 4097–4117. <https://doi.org/10.1111/mec.16755>
- Garud, N. R., Messer, P. W., Buzbas, E. O., & Petrov, D. A. (2015). Recent Selective Sweeps in North American *Drosophila melanogaster* Show Signatures of Soft Sweeps. *PLoS Genetics*, 11(2), e1005004. <https://doi.org/10.1371/journal.pgen.1005004>
- Gerdol, M., Moreira, R., Cruz, F., Gómez-Garrido, J., Vlasova, A., Rosani, U., Venier, P., Naranjo-Ortiz, M. A., Murgarella, M., Greco, S., Balseiro, P., Corvelo, A., Frias, L., Gut, M., Gabaldón, T., Pallavicini, A., Canchaya, C., Novoa, B., Alioto, T. S., ... Figueras, A. (2020). Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biology*, 21(1), 275. <https://doi.org/10.1186/s13059-020-02180-3>
- Gerhold, J. M., Aun, A., Sedman, T., Jöers, P., & Sedman, J. (2010). Strand Invasion Structures in the Inverted Repeat of *Candida albicans* Mitochondrial DNA Reveal a Role for Homologous Recombination in Replication. *Molecular Cell*, 39(6), 851–861. <https://doi.org/10.1016/j.molcel.2010.09.002>
- Giani, A. M., Gallo, G. R., Gianfranceschi, L., & Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal*, 18, 9–19. <https://doi.org/10.1016/j.csbj.2019.11.002>
- Giska, I., Farello, L., Pimenta, J., Seixas, F. A., Ferreira, M. S., Marques, J. P., Miranda, I., Letty, J., Jenny, H., Hackländer, K., Magnussen, E., & Melo-Ferreira, J. (2019). Introgression drives repeated evolution of winter coat color polymorphism in hares. *Proceedings of the National Academy of Sciences*, 116(48), 24150–24156. <https://doi.org/10.1073/pnas.1910471116>
- Giska, I., Pimenta, J., Farello, L., Boursot, P., Hackländer, K., Jenny, H., Reid, N., Montgomery, W. I., Prodöhl, P. A., Alves, P. C., & Melo-Ferreira, J. (2022). The evolutionary pathways for local adaptation in mountain hares. *Molecular Ecology*, 31(5), 1487–1503. <https://doi.org/10.1111/mec.16338>
- Glaserapp, M. R., & Pogson, G. H. (2024). Selection Shapes the Genomic Landscape of Introgressed Ancestry in a Pair of Sympatric Sea Urchin Species. *Genome Biology and Evolution*, 16(6), evae124. <https://doi.org/10.1093/gbe/evae124>
- Goedbloed, D. J., van Hooft, P., Megens, H.-J., Langenbeck, K., Lutz, W., Crooijmans, R. P. M. A., van Wieren, S. E., Ydenberg, R. C., & Prins, H. H. T. (2013). Reintroductions and genetic introgression from domestic pigs have shaped the genetic population structure of Northwest European wild boar. *BMC Genetics*, 14, 43. <https://doi.org/10.1186/1471-2156-14-43>
- Gopalakrishnan, S., Samaniego Castruita, J. A., Sinding, M.-H. S., Kuderna, L. F. K., Räikkönen, J., Petersen, B., Sicheritz-Ponten, T., Larson, G., Orlando, L., Marques-Bonet, T., Hansen, A. J., Dalén, L., & Gilbert, M. T. P. (2017). The wolf reference genome sequence (*Canis lupus lupus*) and its implications for *Canis* spp. Population genomics. *BMC Genomics*, 18(1), 495. <https://doi.org/10.1186/s12864-017-3883-3>
- Grossen, C., Keller, L., Biebach, I., International Goat Genome Consortium, & Croll, D. (2014). Introgression from domestic goat generated variation at the major histocompatibility complex of Alpine ibex. *PLoS Genetics*, 10(6), e1004438. <https://doi.org/10.1371/journal.pgen.1004438>

- Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, *36*(9), 2896–2898. <https://doi.org/10.1093/bioinformatics/btaa025>
- Guan, Y. (2014). Detecting Structure of Haplotypes and Local Ancestry. *Genetics*, *196*(3), 625–642. <https://doi.org/10.1534/genetics.113.160697>
- 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>
- Hagstrom, E., Freyer, C., Battersby, B. J., Stewart, J. B., & Larsson, N.-G. (2014). No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline. *Nucleic Acids Research*, *42*(2), 1111–1116. <https://doi.org/10.1093/nar/gkt969>
- Hamilton, J. A., & Miller, J. M. (2016). Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology*, *30*(1), 33–41. <https://doi.org/10.1111/cobi.12574>
- Harris, T. D., Buzby, P. R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Causey, M., Colonell, J., DiMeo, J., Efcavitch, J. W., Giladi, E., Gill, J., Healy, J., Jarosz, M., Lapen, D., Moulton, K., Quake, S. R., Steinmann, K., Thayer, E., ... Xie, Z. (2008). Single-Molecule DNA Sequencing of a Viral Genome. *Science*, *320*(5872), 106–109. <https://doi.org/10.1126/science.1150427>
- Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques*, *56*(2), 61–77. <https://doi.org/10.2144/000114133>
- Hecht, J., Grewe, F., & Knoop, V. (2011). Extreme RNA Editing in Coding Islands and Abundant Microsatellites in Repeat Sequences of *Selaginella moellendorffii* Mitochondria: The Root of Frequent Plant mtDNA Recombination in Early Tracheophytes. *Genome Biology and Evolution*, *3*, 344–358. <https://doi.org/10.1093/gbe/evr027>
- Hedrick, P. W. (2013). Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, *22*(18), 4606–4618. <https://doi.org/10.1111/mec.12415>
- Heikkinen, M. E., Ruokonen, M., White, T. A., Alexander, M. M., Gündüz, İ., Dobney, K. M., Aspi, J., Searle, J. B., & Pyhäjärvi, T. (2020). Long-Term Reciprocal Gene Flow in Wild and Domestic Geese Reveals Complex Domestication History. *G3 (Bethesda, Md.)*, *10*(9), 3061–3070. <https://doi.org/10.1534/g3.120.400886>
- Hendricks, C. A., & Engelward, B. P. (2004). “Recombomice”: The past, present, and future of recombination–detection in mice. *DNA Repair*, *3*(10), 1255–1261. <https://doi.org/10.1016/j.dnarep.2004.04.006>
- Herbers, E., Kekäläinen, N. J., Hangas, A., Pohjoismäki, J. L., & Goffart, S. (2019). Tissue specific differences in mitochondrial DNA maintenance and expression. *Mitochondrion*, *44*, 85–92. <https://doi.org/10.1016/j.mito.2018.01.004>
- Hermisson, J., & Pennings, P. S. (2017). Soft sweeps and beyond: Understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution*, *8*(6), 700–716. <https://doi.org/10.1111/2041-210X.12808>
- 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>

- Holliday, R. (1964). A mechanism for gene conversion in fungi. *Genetical Research*, 5(2), 282–304. <https://doi.org/10.1017/S0016672300001233>
- Horscroft, C., Ennis, S., Pengelly, R. J., Sluckin, T. J., & Collins, A. (2019). Sequencing era methods for identifying signatures of selection in the genome. *Briefings in Bioinformatics*, 20(6), 1997–2008. <https://doi.org/10.1093/bib/bby064>
- Howe, K., Chow, W., Collins, J., Pelan, S., Pointon, D.-L., Sims, Y., Torrance, J., Tracey, A., & Wood, J. (2021). Significantly improving the quality of genome assemblies through curation. *GigaScience*, 10(1), gaa153. <https://doi.org/10.1093/gigascience/gaa153>
- 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>
- Huang, X., Rymbekova, A., Dolgova, O., Lao, O., & Kuhlwilm, M. (2024). Harnessing deep learning for population genetic inference. *Nature Reviews Genetics*, 25(1), 61–78. <https://doi.org/10.1038/s41576-023-00636-3>
- Huang, Y.-F., Chen, S.-C., Chiang, Y.-S., Chen, T.-H., & Chiu, K.-P. (2012). Palindromic sequence impedes sequencing-by-ligation mechanism. *BMC Systems Biology*, 6(S2), S10. <https://doi.org/10.1186/1752-0509-6-S2-S10>
- Hunter-Zinck, H., & Clark, A. G. (2015). Aberrant Time to Most Recent Common Ancestor as a Signature of Natural Selection. *Molecular Biology and Evolution*, 32(10), 2784–2797. <https://doi.org/10.1093/molbev/msv142>
- Husmeier, D. (2003). Detecting Recombination in 4-Taxa DNA Sequence Alignments with Bayesian Hidden Markov Models and Markov Chain Monte Carlo. *Molecular Biology and Evolution*, 20(3), 315–337. <https://doi.org/10.1093/molbev/msg039>
- Jacobs, G. S., Sluckin, T. J., & Kivisild, T. (2016). Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps. *Genetics*, 203(4), 1807–1825. <https://doi.org/10.1534/genetics.115.185900>
- Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biology*, 17(1), 239. <https://doi.org/10.1186/s13059-016-1103-0>
- James, C., Pemberton, J. M., Navarro, P., & Knott, S. (2022). The impact of SNP density on quantitative genetic analyses of body size traits in a wild population of Soay sheep. *Ecology and Evolution*, 12(12), e9639. <https://doi.org/10.1002/ece3.9639>
- Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science*, 316(5830), 1497–1502. <https://doi.org/10.1126/science.1141319>
- Jones, M. R., Mills, L. S., Alves, P. C., Callahan, C. M., Alves, J. M., Lafferty, D. J. R., Jiggins, F. M., Jensen, J. D., Melo-Ferreira, J., & Good, J. M. (2018). Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science*, 360(6395), 1355–1358. <https://doi.org/10.1126/science.aar5273>
- Jones, M. R., Mills, L. S., Jensen, J. D., & Good, J. M. (2020a). Convergent evolution of seasonal camouflage in response to reduced snow cover across the snowshoe hare range*. *Evolution*, 74(9), 2033–2045. <https://doi.org/10.1111/evo.13976>

- Jones, M. R., Mills, L. S., Jensen, J. D., & Good, J. M. (2020b). The Origin and Spread of Locally Adaptive Seasonal Camouflage in Snowshoe Hares. *The American Naturalist*, *196*(3), 316–332. <https://doi.org/10.1086/710022>
- Kajander, O. A., Karhunen, P. J., Holt, I. J., & Jacobs, H. T. (2001). Prominent mitochondrial DNA recombination intermediates in human heart muscle. *EMBO Reports*, *2*(11), 1007–1012. <https://doi.org/10.1093/embo-reports/kve233>
- Kalinina, A. S., Suvorikova, A. L., Spokoyny, V. G., & Gelfand, M. S. (2016). Detection of homologous recombination in closely related strains. *Journal of Bioinformatics and Computational Biology*, *14*(02), 1641001. <https://doi.org/10.1142/S0219720016410018>
- Karlsson, A.-C., Fallahshahroudi, A., Johnsen, H., Hagenblad, J., Wright, D., Andersson, L., & Jensen, P. (2016). A domestication related mutation in the thyroid stimulating hormone receptor gene (TSHR) modulates photoperiodic response and reproduction in chickens. *General and Comparative Endocrinology*, *228*, 69–78. <https://doi.org/10.1016/j.ygcen.2016.02.010>
- Kaye, A. M., & Wasserman, W. W. (2021). The genome atlas: Navigating a new era of reference genomes. *Trends in Genetics*, *37*(9), 807–818. <https://doi.org/10.1016/j.tig.2020.12.002>
- Kazak, L., Reyes, A., & Holt, I. J. (2012). Minimizing the damage: Repair pathways keep mitochondrial DNA intact. *Nature Reviews Molecular Cell Biology*, *13*(10), 659–671. <https://doi.org/10.1038/nrm3439>
- Ke, C. B., He, W. S., Li, C. J., Shi, D., Gao, F., & Tian, Y. K. (2012). Enhanced SCN7A/Nax expression contributes to bone cancer pain by increasing excitability of neurons in dorsal root ganglion. *Neuroscience*, *227*, 80–89. <https://doi.org/10.1016/j.neuroscience.2012.09.046>
- Kejnovsky, E., Hobza, R., Cermak, T., Kubat, Z., & Vyskot, B. (2009). The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity*, *102*(6), 533–541. <https://doi.org/10.1038/hdy.2009.17>
- Kemper, K. E., Saxton, S. J., Bolormaa, S., Hayes, B. J., & Goddard, M. E. (2014). Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics*, *15*(1), 246. <https://doi.org/10.1186/1471-2164-15-246>
- Kerpedjiev, P., Abdennur, N., Lekschas, F., McCallum, C., Dinkla, K., Strobel, H., Lubner, J. M., Ouellette, S. B., Azhir, A., Kumar, N., Hwang, J., Lee, S., Alver, B. H., Pfister, H., Mirny, L. A., Park, P. J., & Gehlenborg, N. (2018). HiGlass: Web-based visual exploration and analysis of genome interaction maps. *Genome Biology*, *19*(1), 125. <https://doi.org/10.1186/s13059-018-1486-1>
- Kim, Y., & Nielsen, R. (2004). Linkage Disequilibrium as a Signature of Selective Sweeps. *Genetics*, *167*(3), 1513–1524. <https://doi.org/10.1534/genetics.103.025387>
- Kim, Y., & Stephan, W. (2002). Detecting a Local Signature of Genetic Hitchhiking Along a Recombining Chromosome. *Genetics*, *160*(2), 765–777. <https://doi.org/10.1093/genetics/160.2.765>
- Kingan, S. B., Heaton, H., Cudini, J., Lambert, C. C., Baybayan, P., Galvin, B. D., Durbin, R., Korlach, J., & Lawniczak, M. K. N. (2019). A High-Quality De novo Genome Assembly from a Single Mosquito Using PacBio Sequencing. *Genes*, *10*(1), 62. <https://doi.org/10.3390/genes10010062>
- Kinsella, R. J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., Kersey, P., & Flicek, P. (2011). Ensembl

- BioMarts: A hub for data retrieval across taxonomic space. *Database: The Journal of Biological Databases and Curation*, 2011, bar030. <https://doi.org/10.1093/database/bar030>
- Kolmogorov, M., Billingsley, K. J., Mastoras, M., Meredith, M., Monlong, J., Lorig-Roach, R., Asri, M., Alvarez Jerez, P., Malik, L., Dewan, R., Reed, X., Genner, R. M., Daida, K., Behera, S., Shafin, K., Pesout, T., Prabakaran, J., Carnevali, P., Yang, J., ... Paten, B. (2023). Scalable Nanopore sequencing of human genomes provides a comprehensive view of haplotype-resolved variation and methylation. *Nature Methods*, 20(10), 1483–1492. <https://doi.org/10.1038/s41592-023-01993-x>
- Kong, S., & Kubatko, L. S. (2021). Comparative Performance of Popular Methods for Hybrid Detection using Genomic Data. *Systematic Biology*, 70(5), 891–907. <https://doi.org/10.1093/sysbio/syaa092>
- Kornblum, C., Nicholls, T. J., Haack, T. B., Schöler, S., Peeva, V., Danhauser, K., Hallmann, K., Zsurka, G., Rorbach, J., Iuso, A., Wieland, T., Sciacco, M., Ronchi, D., Comi, G. P., Moggio, M., Quinzii, C. M., DiMauro, S., Calvo, S. E., Mootha, V. K., ... Prokisch, H. (2013). Loss-of-function mutations in MGME1 impair mtDNA replication and cause multisystemic mitochondrial disease. *Nature Genetics*, 45(2), 214–219. <https://doi.org/10.1038/ng.2501>
- Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. W. (2006). GARD: A genetic algorithm for recombination detection. *Bioinformatics*, 22(24), 3096–3098. <https://doi.org/10.1093/bioinformatics/btl474>
- Kozik, A., Rowan, B. A., Lavelle, D., Berke, L., Schranz, M. E., Michelmore, R. W., & Christensen, A. C. (2019). The alternative reality of plant mitochondrial DNA: One ring does not rule them all. *PLOS Genetics*, 15(8), e1008373. <https://doi.org/10.1371/journal.pgen.1008373>
- Kramara, J., Osia, B., & Malkova, A. (2018). Break-Induced Replication: The Where, The Why, and The How. *Trends in Genetics: TIG*, 34(7), 518–531. <https://doi.org/10.1016/j.tig.2018.04.002>
- Kraytsberg, Y., Schwartz, M., Brown, T. A., Ebraldise, 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>
- Krutikova, A., Dementieva, N., Shcherbakov, Y., Goncharov, V., Griffin, D., & Romanov, M. (2024). Polymorphism of Genes Potentially Affecting Growth and Body Size Suggests Genetic Divergence in Wild and Domestic Reindeer (*Rangifer tarandus*) Populations. *Genes*, 15(12), 1629. <https://doi.org/10.3390/genes15121629>
- Kubatko, L. S., & Chifman, J. (2019). An invariants-based method for efficient identification of hybrid species from large-scale genomic data. *BMC Evolutionary Biology*, 19(1), 112. <https://doi.org/10.1186/s12862-019-1439-7>
- Kyriakidou, M., Tai, H. H., Anglin, N. L., Ellis, D., & Strömvik, M. V. (2018). Current Strategies of Polyploid Plant Genome Sequence Assembly. *Frontiers in Plant Science*, 9, 1660. <https://doi.org/10.3389/fpls.2018.01660>
- Ladoukakis, E. D., Theologidis, I., Rodakis, G. C., & Zouros, E. (2011). Homologous Recombination between Highly Diverged Mitochondrial Sequences: Examples from Maternally and Paternally Transmitted Genomes. *Molecular Biology and Evolution*, 28(6), 1847–1859. <https://doi.org/10.1093/molbev/msr007>

- 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>
- Ladoukakis, E. D., & Zouros, E. (2017). Evolution and inheritance of animal mitochondrial DNA: Rules and exceptions. *Journal of Biological Research-Thessaloniki*, *24*(1), 2. <https://doi.org/10.1186/s40709-017-0060-4>
- Lange, J. D., & Pool, J. E. (2016). A haplotype method detects diverse scenarios of local adaptation from genomic sequence variation. *Molecular Ecology*, *25*(13), 3081–3100. <https://doi.org/10.1111/mec.13671>
- Lawal, R. A., & Hanotte, O. (2021). Domestic chicken diversity: Origin, distribution, and adaptation. *Animal Genetics*, *52*(4), 385–394. <https://doi.org/10.1111/age.13091>
- Lawniczak, M. K. N., Durbin, R., Flicek, P., Lindblad-Toh, K., Wei, X., Archibald, J. M., Baker, W. J., Belov, K., Blaxter, M. L., Marques Bonet, T., Childers, A. K., Coddington, J. A., Crandall, K. A., Crawford, A. J., Davey, R. P., Di Palma, F., Fang, Q., Haerty, W., Hall, N., ... Richards, S. (2022). Standards recommendations for the Earth BioGenome Project. *Proceedings of the National Academy of Sciences*, *119*(4), e2115639118. <https://doi.org/10.1073/pnas.2115639118>
- 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>
- Levänen, R., Kunnasranta, M., & Pohjoismäki, J. (2018a). Mitochondrial DNA Introgression at the Northern Edge of the Brown Hare (*Lepus europaeus*) Range. *Annales Zoologici Fennici*, *55*(1–3), 15–24. <https://doi.org/10.5735/086.055.0103>
- 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*(1), e0191790. <https://doi.org/10.1371/journal.pone.0191790>
- Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., Durbin, R., Edwards, S. V., Forest, F., Gilbert, M. T. P., Goldstein, M. M., Grigoriev, I. V., Hackett, K. J., Haussler, D., Jarvis, E. D., Johnson, W. E., Patrinos, A., Richards, S., Castilla-Rubio, J. C., ... Zhang, G. (2018). Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences*, *115*(17), 4325–4333. <https://doi.org/10.1073/pnas.1720115115>
- Li, C., Li, Y., Zheng, J., Guo, Z., Mei, X., Lei, M., Ren, Y., Zhang, X., Zhang, C., Yang, C., Tang, L., Ji, Y., Yang, R., Yu, J., Xie, X., & Kuang, L. (2022). Trait Analysis in Domestic Rabbits (*Oryctolagus cuniculus* f. *Domesticus*) Using SNP Markers from Genotyping-by-Sequencing Data. *Animals: An Open Access Journal from MDPI*, *12*(16), 2052. <https://doi.org/10.3390/ani12162052>
- Li, D., Roca, M., Yuecel, R., & Lorenz, A. (2019). Immediate visualization of recombination events and chromosome segregation defects in fission yeast meiosis. *Chromosoma*, *128*(3), 385–396. <https://doi.org/10.1007/s00412-019-00691-y>
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, H., Dawood, M., Khayat, M. M., Farek, J. R., Jhangiani, S. N., Khan, Z. M., Mitani, T., Coban-Akdemir, Z., Lupski, J. R., Venner, E., Posey, J. E., Sabo, A., & Gibbs, R. A. (2021).

- Exome variant discrepancies due to reference-genome differences. *The American Journal of Human Genetics*, 108(7), 1239–1250. <https://doi.org/10.1016/j.ajhg.2021.05.011>
- Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496. <https://doi.org/10.1038/nature10231>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liao, W.-W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., Lu, S., Lucas, J. K., Monlong, J., Abel, H. J., Buonaiuto, S., Chang, X. H., Cheng, H., Chu, J., Colonna, V., Eizenga, J. M., Feng, X., Fischer, C., Fulton, R. S., ... Paten, B. (2023). A draft human pangenome reference. *Nature*, 617(7960), 312–324. <https://doi.org/10.1038/s41586-023-05896-x>
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragozy, 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 (New York, N.Y.)*, 326(5950), 289–293. <https://doi.org/10.1126/science.1181369>
- Ling, F. (2002). Recombination-dependent mtDNA partitioning: In vivo role of Mhr1p to promote pairing of homologous DNA. *The EMBO Journal*, 21(17), 4730–4740. <https://doi.org/10.1093/emboj/cdf466>
- Linsmith, G., Rombauts, S., Montanari, S., Deng, C. H., Celton, J.-M., Guérif, P., Liu, C., Lohaus, R., Zurn, J. D., Cestaro, A., Bassil, N. V., Bakker, L. V., Schijlen, E., Gardiner, S. E., Lespinasse, Y., Durel, C.-E., Velasco, R., Neale, D. B., Chagné, D., ... Bianco, L. (2019). Pseudo-chromosome-length genome assembly of a double haploid “Bartlett” pear (*Pyrus communis* L.). *GigaScience*, 8(12), giz138. <https://doi.org/10.1093/gigascience/giz138>
- Lister, R., O’Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., & Ecker, J. R. (2008). Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell*, 133(3), 523–536. <https://doi.org/10.1016/j.cell.2008.03.029>
- Liu, C., Wang, S., Dong, X., Zhao, J., Ye, X., Gong, R., & Ren, Z. (2021). Exploring the genomic resources and analysing the genetic diversity and population structure of Chinese indigenous rabbit breeds by RAD-seq. *BMC Genomics*, 22(1), 573. <https://doi.org/10.1186/s12864-021-07833-6>
- Liu, Y., & Xie, J. (2020). Cauchy Combination Test: A Powerful Test With Analytic p -Value Calculation Under Arbitrary Dependency Structures. *Journal of the American Statistical Association*, 115(529), 393–402. <https://doi.org/10.1080/01621459.2018.1554485>
- Lloret-Villas, A., Bhati, M., Kadri, N. K., Fries, R., & Pausch, H. (2021). Investigating the impact of reference assembly choice on genomic analyses in a cattle breed. *BMC Genomics*, 22(1), 363. <https://doi.org/10.1186/s12864-021-07554-w>
- Lloveras, L., Maroto, J., Soler, J., Thomas, R., Moreno-García, M., Nadal, J., & Soler, N. (2016). The role of small prey in human subsistence strategies from Early Upper Palaeolithic sites in Iberia: The rabbits from the Evolved Aurignacian level of Arbreda Cave. *Journal of Quaternary Science*, 31(5), 458–471. <https://doi.org/10.1002/jqs.2869>

- Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., & Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, *30*(5), 434–439. <https://doi.org/10.1038/nbt.2198>
- Lu, Y., Li, M., Gao, Z., Ma, H., Chong, Y., Hong, J., Wu, J., Wu, D., Xi, D., & Deng, W. (2025). Advances in Whole Genome Sequencing: Methods, Tools, and Applications in Population Genomics. *International Journal of Molecular Sciences*, *26*(1), 372. <https://doi.org/10.3390/ijms26010372>
- Lunt, D. H., Whipple, L. E., & Hyman, B. C. (1998). Mitochondrial DNA variable number tandem repeats (VNTRs): Utility and problems in molecular ecology. *Molecular Ecology*, *7*(11), 1441–1455. <https://doi.org/10.1046/j.1365-294x.1998.00495.x>
- Lynch, M., Bost, D., Wilson, S., Maruki, T., & Harrison, S. (2014). Population-genetic inference from pooled-sequencing data. *Genome Biology and Evolution*, *6*(5), 1210–1218. <https://doi.org/10.1093/gbe/evu085>
- Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q., & Simianer, H. (2015). Properties of different selection signature statistics and a new strategy for combining them. *Heredity*, *115*(5), 426–436. <https://doi.org/10.1038/hdy.2015.42>
- Makvandi-Nejad, S., Hoffman, G. E., Allen, J. J., Chu, E., Gu, E., Chandler, A. M., Loredó, A. I., Bellone, R. R., Mezey, J. G., Brooks, S. A., & Sutter, N. B. (2012). Four Loci Explain 83% of Size Variation in the Horse. *PLoS ONE*, *7*(7), e39929. <https://doi.org/10.1371/journal.pone.0039929>
- Manekar, S. C., & Sathe, S. R. (2019). Estimating the k-mer Coverage Frequencies in Genomic Datasets: A Comparative Assessment of the State-of-the-art. *Current Genomics*, *20*(1), 2–15. <https://doi.org/10.2174/1389202919666181026101326>
- 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>
- Mapleson, D., Garcia Accinelli, G., Kettleborough, G., Wright, J., & Clavijo, B. J. (2017). KAT: A K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics (Oxford, England)*, *33*(4), 574–576. <https://doi.org/10.1093/bioinformatics/btw663>
- Maréchal, A., & Brisson, N. (2010). Recombination and the maintenance of plant organelle genome stability. *New Phytologist*, *186*(2), 299–317. <https://doi.org/10.1111/j.1469-8137.2010.03195.x>
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., ... Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*(7057), 376–380. <https://doi.org/10.1038/nature03959>
- Marín-García, P. J., & Llobat, L. (2021). What Are the Keys to the Adaptive Success of European Wild Rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula? *Animals: An Open Access Journal from MDPI*, *11*(8), 2453. <https://doi.org/10.3390/ani11082453>
- 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>

- Martin, D. P., Murrell, B., Khoosal, A., & Muhire, B. (2017). Detecting and Analyzing Genetic Recombination Using RDP4. In J. M. Keith (Ed.), *Bioinformatics* (Vol. 1525, pp. 433–460). Springer New York. https://doi.org/10.1007/978-1-4939-6622-6_17
- Martin, D., & Rybicki, E. (2000). RDP: Detection of recombination amongst aligned sequences. *Bioinformatics*, *16*(6), 562–563. <https://doi.org/10.1093/bioinformatics/16.6.562>
- Martin, F. J., Amode, M. R., Aneja, A., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S. K., Bignell, A., Boddu, S., Branco Lins, P. R., Brooks, L., Ramaraju, S. B., Charkhchi, M., Cockburn, A., Da Rin Fiorretto, L., ... Flicek, P. (2023). Ensembl 2023. *Nucleic Acids Research*, *51*(D1), D933–D941. <https://doi.org/10.1093/nar/gkac958>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. <https://doi.org/10.14806/ej.17.1.200>
- Mattucci, F., Galaverni, M., Lyons, L. A., Alves, P. C., Randi, E., Velli, E., Pagani, L., & Caniglia, R. (2019). Genomic approaches to identify hybrids and estimate admixture times in European wildcat populations. *Scientific Reports*, *9*(1), 11612. <https://doi.org/10.1038/s41598-019-48002-w>
- Melo-Ferreira, J., Alves, P. C., Freitas, H., Ferrand, N., & Boursot, P. (2009). The genomic legacy from the extinct *Lepus timidus* to the three hare species of Iberia: Contrast between mtDNA, sex chromosomes and autosomes. *Molecular Ecology*, *18*(12), 2643–2658. <https://doi.org/10.1111/j.1365-294X.2009.04221.x>
- 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>
- Messing, J., Crea, R., & Seeburg, P. H. (1981). A system for shotgun DNA sequencing. *Nucleic Acids Research*, *9*(2), 309–321. <https://doi.org/10.1093/nar/9.2.309>
- Metzker, M. L., Lu, J., & Gibbs, R. A. (1996). Electrophoretically Uniform Fluorescent Dyes for Automated DNA Sequencing. *Science*, *271*(5254), 1420–1422. <https://doi.org/10.1126/science.271.5254.1420>
- 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>
- Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., & Gurevich, A. (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, *34*(13), i142–i150. <https://doi.org/10.1093/bioinformatics/bty266>
- Milián-García, Y., Ramos-Targarona, R., Pérez-Fleitas, E., Sosa-Rodríguez, G., Guerra-Manchena, L., Alonso-Tabet, M., Espinosa-López, G., & Russello, M. A. (2015). Genetic evidence of hybridization between the critically endangered Cuban crocodile and the American crocodile: Implications for population history and in situ/ex situ conservation. *Heredity*, *114*(3), 272–280. <https://doi.org/10.1038/hdy.2014.96>
- Minoche, A. E., Dohm, J. C., & Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biology*, *12*(11), R112. <https://doi.org/10.1186/gb-2011-12-11-r112>
- Miura, T., Yamana, Y., Usui, T., Ogawa, H. I., Yamamoto, M.-T., & Kusano, K. (2012). Homologous recombination via synthesis-dependent strand annealing in yeast requires the

- Irc20 and Srs2 DNA helicases. *Genetics*, 191(1), 65–78. <https://doi.org/10.1534/genetics.112.139105>
- Montaña-Lozano, P., Moreno-Carmona, M., Ochoa-Capera, M., Medina, N. S., Boore, J. L., & Prada, C. F. (2022). Comparative genomic analysis of vertebrate mitochondrial reveals a differential of rearrangements rate between taxonomic class. *Scientific Reports*, 12(1), 5479. <https://doi.org/10.1038/s41598-022-09512-2>
- Montero-Tena, J. A., Abdollahi Sisi, N., Kox, T., Abbadi, A., Snowdon, R. J., & Golicz, A. A. (2024). haploMAGIC: Accurate phasing and detection of recombination in multiparental populations despite genotyping errors. *G3 (Bethesda, Md.)*, 14(8), jkae109. <https://doi.org/10.1093/g3journal/jkae109>
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, 51(0), 263–273. <https://doi.org/10.1101/SQB.1986.051.01.032>
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science*, 320(5881), 1344–1349. <https://doi.org/10.1126/science.1158441>
- Navarro, A., & Barton, N. H. (2003). ACCUMULATING POSTZYGOTIC ISOLATION GENES IN PARAPATRY: A NEW TWIST ON CHROMOSOMAL SPECIATION. *Evolution*, 57(3), 447–459. <https://doi.org/10.1111/j.0014-3820.2003.tb01537.x>
- 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>
- Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E., Bamshad, M., Nickerson, D. A., & Shendure, J. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 461(7261), 272–276. <https://doi.org/10.1038/nature08250>
- Nicholls, T. J., Zsurka, G., Peeva, V., Schöler, S., Szczesny, R. J., Cysewski, D., Reyes, A., Kornblum, C., Sciacco, M., Moggio, M., Dziembowski, A., Kunz, W. S., & Minczuk, M. (2014). Linear mtDNA fragments and unusual mtDNA rearrangements associated with pathological deficiency of MGME1 exonuclease. *Human Molecular Genetics*, 23(23), 6147–6162. <https://doi.org/10.1093/hmg/ddu336>
- Nielsen, R. (2005). Molecular Signatures of Natural Selection. *Annual Review of Genetics*, 39(1), 197–218. <https://doi.org/10.1146/annurev.genet.39.073003.112420>
- Niimi, M., Chen, Y., Yan, H., Wang, Y., Koike, T., & Fan, J. (2020). Hyperlipidemic Rabbit Models for Anti-Atherosclerotic Drug Development. *Applied Sciences*, 10(23), 8681. <https://doi.org/10.3390/app10238681>
- Oliveira, R., Godinho, R., Randi, E., & Alves, P. C. (2008). Hybridization versus conservation: Are domestic cats threatening the genetic integrity of wildcats (*Felis silvestris silvestris*) in Iberian Peninsula? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 363(1505), 2953–2961. <https://doi.org/10.1098/rstb.2008.0052>
- Panigrahi, M., Rajawat, D., Nayak, S. S., Ghildiyal, K., Sharma, A., Jain, K., Lei, C., Bhushan, B., Mishra, B. P., & Dutt, T. (2023). Landmarks in the history of selective sweeps. *Animal Genetics*, 54(6), 667–688. <https://doi.org/10.1111/age.13355>

- Papachristou, D., Koutsouli, P., Laliotis, G. P., Kunz, E., Upadhyay, M., Seichter, D., Russ, I., Gjoko, B., Kostaras, N., Bizelis, I., & Medugorac, I. (2020). Genomic diversity and population structure of the indigenous Greek and Cypriot cattle populations. *Genetics, Selection, Evolution: GSE*, 52(1), 43. <https://doi.org/10.1186/s12711-020-00560-8>
- Paten, B., Novak, A. M., Eizenga, J. M., & Garrison, E. (2017). Genome graphs and the evolution of genome inference. *Genome Research*, 27(5), 665–676. <https://doi.org/10.1101/gr.214155.116>
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, 192(3), 1065–1093. <https://doi.org/10.1534/genetics.112.145037>
- Patterson, N., Price, A. L., & Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genetics*, 2(12), e190. <https://doi.org/10.1371/journal.pgen.0020190>
- Paul, P., Nag, D., & Chakraborty, S. (2016). Recombination hotspots: Models and tools for detection. *DNA Repair*, 40, 47–56. <https://doi.org/10.1016/j.dnarep.2016.02.005>
- Pavlidis, P., & Alachiotis, N. (2017). A survey of methods and tools to detect recent and strong positive selection. *Journal of Biological Research-Thessaloniki*, 24(1), 7. <https://doi.org/10.1186/s40709-017-0064-0>
- Pavlidis, P., Živković, D., Stamatakis, A., & Alachiotis, N. (2013). SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. *Molecular Biology and Evolution*, 30(9), 2224–2234. <https://doi.org/10.1093/molbev/mst112>
- Payne, A., Holmes, N., Rakyant, V., & Loose, M. (2019). BulkVis: A graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics*, 35(13), 2193–2198. <https://doi.org/10.1093/bioinformatics/bty841>
- Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, 25(11), 2337–2360. <https://doi.org/10.1111/mec.13557>
- Peeva, V., Blei, D., Trombly, G., Corsi, S., Szukszto, M. J., Rebelo-Guimar, 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>
- Peischl, S., Kirkpatrick, M., & Excoffier, L. (2015). Expansion Load and the Evolutionary Dynamics of a Species Range. *The American Naturalist*, 185(4), E81–E93. <https://doi.org/10.1086/680220>
- Peñalba, J. V., & Wolf, J. B. W. (2020). From molecules to populations: Appreciating and estimating recombination rate variation. *Nature Reviews Genetics*, 21(8), 476–492. <https://doi.org/10.1038/s41576-020-0240-1>
- Pennings, P. S., & Hermisson, J. (2006). Soft Sweeps III: The Signature of Positive Selection from Recurrent Mutation. *PLoS Genetics*, 2(12), e186. <https://doi.org/10.1371/journal.pgen.0020186>
- Peona, V., Weissensteiner, M. H., & Suh, A. (2018). How complete are “complete” genome assemblies?—An avian perspective. *Molecular Ecology Resources*, 18(6), 1188–1195. <https://doi.org/10.1111/1755-0998.12933>
- Persson, Ö., Muthukumar, Y., Basu, S., Jenninger, L., Uhler, J. P., Berglund, A.-K., McFarland, R., Taylor, R. W., Gustafsson, C. M., Larsson, E., & Falkenberg, M. (2019). Copy-choice

- recombination during mitochondrial L-strand synthesis causes DNA deletions. *Nature Communications*, 10(1), 759. <https://doi.org/10.1038/s41467-019-08673-5>
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, 7(5), e37135. <https://doi.org/10.1371/journal.pone.0037135>
- Petr, M., Vernot, B., & Kelso, J. (2019). Admixr-R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics (Oxford, England)*, 35(17), 3194–3195. <https://doi.org/10.1093/bioinformatics/btz030>
- Phan, L., Zhang, H., Wang, Q., Villamarin, R., Hefferon, T., Ramanathan, A., & Kattman, B. (2025). The evolution of dbSNP: 25 years of impact in genomic research. *Nucleic Acids Research*, 53(D1), D925–D931. <https://doi.org/10.1093/nar/gkae977>
- Phillips, A. F., Millet, A. R., Tigano, M., Dubois, S. M., Crimmins, H., Babin, L., Charpentier, M., Piganeau, M., Brunet, E., & Sfeir, A. (2017). Single-Molecule Analysis of mtDNA Replication Uncovers the Basis of the Common Deletion. *Molecular Cell*, 65(3), 527–538.e6. <https://doi.org/10.1016/j.molcel.2016.12.014>
- 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>
- Pilot, M., Greco, C., vonHoldt, B. M., Randi, E., Jędrzejewski, W., Sidorovich, V. E., Konopiński, M. K., Ostrander, E. A., & Wayne, R. K. (2018). Widespread, long-term admixture between grey wolves and domestic dogs across Eurasia and its implications for the conservation status of hybrids. *Evolutionary Applications*, 11(5), 662–680. <https://doi.org/10.1111/eva.12595>
- Pilot, M., Moura, A. E., Okhlopkov, I. M., Mamaev, N. V., Manaseryan, N. H., Hayrapetyan, V., Kopaliani, N., Tsingarska, E., Alagaili, A. N., Mohammed, O. B., Ostrander, E. A., & Bogdanowicz, W. (2021). Human-modified canids in human-modified landscapes: The evolutionary consequences of hybridization for grey wolves and free-ranging domestic dogs. *Evolutionary Applications*, 14(10), 2433–2456. <https://doi.org/10.1111/eva.13257>
- Pohjoismäki, J. L. O., Goffart, S., Tyynismaa, H., Willcox, S., Ide, T., Kang, D., Suomalainen, A., Karhunen, P. J., Griffith, J. D., Holt, I. J., & Jacobs, H. T. (2009). Human Heart Mitochondrial DNA Is Organized in Complex Catenated Networks Containing Abundant Four-way Junctions and Replication Forks. *Journal of Biological Chemistry*, 284(32), 21446–21457. <https://doi.org/10.1074/jbc.M109.016600>
- Pohjoismäki, J. L. O., Michell, C., Levänen, R., & Smith, S. (2021). Hybridization with mountain hares increases the functional allelic repertoire in brown hares. *Scientific Reports*, 11(1), 15771. <https://doi.org/10.1038/s41598-021-95357-0>
- Pohjoismäki, J. L. O., Williams, S. L., Boettger, T., Goffart, S., Kim, J., Suomalainen, A., Moraes, C. T., & Braun, T. (2013). Overexpression of Twinkle-helicase protects cardiomyocytes from genotoxic stress caused by reactive oxygen species. *Proceedings of the National Academy of Sciences*, 110(48), 19408–19413. <https://doi.org/10.1073/pnas.1303046110>
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8), 904–909. <https://doi.org/10.1038/ng1847>

- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Provine, W. B. (2004). Ernst Mayr. *Genetics*, *167*(3), 1041–1046. <https://doi.org/10.1093/genetics/167.3.1041>
- Pybus, M., Luisi, P., Dall’Olio, G. M., Uzkuđun, M., Laayouni, H., Bertranpetit, J., & Engelken, J. (2015). Hierarchical boosting: A machine-learning framework to detect and classify hard selective sweeps in human populations. *Bioinformatics*, *31*(24), 3946–3952. <https://doi.org/10.1093/bioinformatics/btv493>
- Quilodrán, C. S., Nussberger, B., Macdonald, D. W., Montoya-Burgos, J. I., & Currat, M. (2020). Projecting introgression from domestic cats into European wildcats in the Swiss Jura. *Evolutionary Applications*, *13*(8), 2101–2112. <https://doi.org/10.1111/eva.12968>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Radman, M., Matic, I., & Taddei, F. (1999). Evolution of Evolvability^a. *Annals of the New York Academy of Sciences*, *870*(1), 146–155. <https://doi.org/10.1111/j.1749-6632.1999.tb08874.x>
- Ramos-Onsins, S. E., Burgos-Paz, W., Manunza, A., & Amills, M. (2014). Mining the pig genome to investigate the domestication process. *Heredity*, *113*(6), 471–484. <https://doi.org/10.1038/hdy.2014.68>
- Randhawa, I. A., Khatkar, M., Thomson, P., & Raadsma, H. (2014). Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. *BMC Genetics*, *15*(1), 34. <https://doi.org/10.1186/1471-2156-15-34>
- Rees, D. C., Williams, T. N., & Gladwin, M. T. (2010). Sick cell disease. *The Lancet*, *376*(9757), 2018–2031. [https://doi.org/10.1016/S0140-6736\(10\)61029-X](https://doi.org/10.1016/S0140-6736(10)61029-X)
- 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. (2020a). *Towards complete and error-free genome assemblies of all vertebrate species*. <https://doi.org/10.1101/2020.05.22.110833>
- Rhie, A., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020b). Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, *21*(1), 245. <https://doi.org/10.1186/s13059-020-02134-9>
- Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and its Applications. *Genomics, Proteomics & Bioinformatics*, *13*(5), 278–289. <https://doi.org/10.1016/j.gpb.2015.08.002>
- Rhymer, J. M., & Simberloff, D. (1996). EXTINCTION BY HYBRIDIZATION AND INTROGRESSION. *Annual Review of Ecology and Systematics*, *27*(1), 83–109. <https://doi.org/10.1146/annurev.ecolsys.27.1.83>
- Robinson, N. P. (2013). Analysis of Branched DNA Replication and Recombination Intermediates from Prokaryotic Cells by Two-Dimensional (2D) Native–Native Agarose Gel Electrophoresis. In S. Makovets (Ed.), *DNA Electrophoresis* (Vol. 1054, pp. 45–61). Humana Press. https://doi.org/10.1007/978-1-62703-565-1_3
- Rochus, C. M., Tortereau, F., Plisson-Petit, F., Restoux, G., Moreno-Romieux, C., Tossier-Klopp, G., & Servin, B. (2018). Revealing the selection history of adaptive loci using genome-

- wide scans for selection: An example from domestic sheep. *BMC Genomics*, 19(1), 71. <https://doi.org/10.1186/s12864-018-4447-x>
- Rostamzadeh Mahdabi, E., Esmailzadeh, A., Ayatollahi Mehrjadi, A., & Asadi Fozi, M. (2021). A genome-wide scan to identify signatures of selection in two Iranian indigenous chicken ecotypes. *Genetics, Selection, Evolution: GSE*, 53(1), 72. <https://doi.org/10.1186/s12711-021-00664-9>
- Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., ... Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475(7356), 348–352. <https://doi.org/10.1038/nature10242>
- Saif, R., Henkel, J., Jagannathan, V., Drögemüller, C., Flury, C., & Leeb, T. (2020). The LCORL Locus Is under Selection in Large-Sized Pakistani Goat Breeds. *Genes*, 11(2), 168. <https://doi.org/10.3390/genes11020168>
- Sambrook, J., & Russell, D. W. (2006). Purification of Nucleic Acids by Extraction with Phenol:Chloroform. *Cold Spring Harbor Protocols*, 2006(1), pdb.prot4455. <https://doi.org/10.1101/pdb.prot4455>
- Sancho, S., Moraes, C. T., Tanji, K., & Miranda, A. F. (1992). Structural and functional mitochondrial abnormalities associated with high levels of partially deleted mitochondrial DNAs in somatic cell hybrids. *Somatic Cell and Molecular Genetics*, 18(5), 431–442. <https://doi.org/10.1007/BF01233083>
- Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94(3), 441–448. [https://doi.org/10.1016/0022-2836\(75\)90213-2](https://doi.org/10.1016/0022-2836(75)90213-2)
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
- Sankararaman, S., Kimmel, G., Halperin, E., & Jordan, M. I. (2008). On the inference of ancestries in admixed populations. *Genome Research*, 18(4), 668–675. <https://doi.org/10.1101/gr.072751.107>
- Sardell, J. M., & Kirkpatrick, M. (2020). Sex Differences in the Recombination Landscape. *The American Naturalist*, 195(2), 361–379. <https://doi.org/10.1086/704943>
- Sato, A., Nakada, K., Akimoto, M., Ishikawa, K., Ono, T., Shitara, H., Yonekawa, H., & Hayashi, J.-I. (2005). Rare creation of recombinant mtDNA haplotypes in mammalian tissues. *Proceedings of the National Academy of Sciences*, 102(17), 6057–6062. <https://doi.org/10.1073/pnas.0408666102>
- Sato, D. X., Rafati, N., Ring, H., Younis, S., Feng, C., Blanco-Aguiar, J. A., Rubin, C.-J., Villafuerte, R., Hallböök, F., Carneiro, M., & Andersson, L. (2020). Brain Transcriptomics of Wild and Domestic Rabbits Suggests That Changes in Dopamine Signaling and Ciliary Function Contributed to Evolution of Tameness. *Genome Biology and Evolution*, 12(10), 1918–1928. <https://doi.org/10.1093/gbe/evaa158>
- Saunders, I. W., Brohede, J., & Hannan, G. N. (2007). Estimating genotyping error rates from Mendelian errors in SNP array genotypes and their impact on inference. *Genomics*, 90(3), 291–296. <https://doi.org/10.1016/j.ygeno.2007.05.011>

- Sawyer, N. T., Helvig, A. W., Makinson, C. D., Decker, M. J., Neigh, G. N., & Escayg, A. (2016). *Scn1a* dysfunction alters behavior but not the effect of stress on seizure response. *Genes, Brain and Behavior*, *15*(3), 335–347. <https://doi.org/10.1111/gbb.12281>
- Schrider, D. R., & Kern, A. D. (2016). S/HIC: Robust Identification of Soft and Hard Sweeps Using Machine Learning. *PLOS Genetics*, *12*(3), e1005928. <https://doi.org/10.1371/journal.pgen.1005928>
- Sereika, M., Kirkegaard, R. H., Karst, S. M., Michaelsen, T. Y., Sørensen, E. A., Wollenberg, R. D., & Albertsen, M. (2022). Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nature Methods*, *19*(7), 823–826. <https://doi.org/10.1038/s41592-022-01539-7>
- Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. W. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, *8*(8), 907–917. <https://doi.org/10.1111/2041-210X.12700>
- Shedge, V., Arrieta-Montiel, M., Christensen, A. C., & Mackenzie, S. A. (2007). Plant Mitochondrial Recombination Surveillance Requires Unusual *RecA* and *MutS* Homologs. *The Plant Cell*, *19*(4), 1251–1264. <https://doi.org/10.1105/tpc.106.048355>
- Sheehan, S., & Song, Y. S. (2016). Deep Learning for Population Genetic Inference. *PLOS Computational Biology*, *12*(3), e1004845. <https://doi.org/10.1371/journal.pcbi.1004845>
- 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>
- Shokolenko, I. N., Wilson, G. L., & Alexeyev, M. F. (2013). Persistent damage induces mitochondrial DNA degradation. *DNA Repair*, *12*(7), 488–499. <https://doi.org/10.1016/j.dnarep.2013.04.023>
- Signer-Hasler, H., Henkel, J., Bangerter, E., Bulut, Z., The VarGoats Consortium, Drögemüller, C., Leeb, T., & Flury, C. (2022). Runs of homozygosity in Swiss goats reveal genetic changes associated with domestication and modern selection. *Genetics Selection Evolution*, *54*(1), 6. <https://doi.org/10.1186/s12711-022-00695-w>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Simpson, J. T. (2014). Exploring genome characteristics and sequence quality without a reference. *Bioinformatics*, *30*(9), 1228–1235. <https://doi.org/10.1093/bioinformatics/btu023>
- Singh, N. A., Pappas, C., Dahle, E. J., Claes, L. R. F., Pruess, T. H., De Jonghe, P., Thompson, J., Dixon, M., Gurnett, C., Peiffer, A., White, H. S., Filloux, F., & Leppert, M. F. (2009). A Role of SCN9A in Human Epilepsies, As a Cause of Febrile Seizures and As a Potential Modifier of Dravet Syndrome. *PLoS Genetics*, *5*(9), e1000649. <https://doi.org/10.1371/journal.pgen.1000649>
- Smit, A., Hubley, R., & Green, P. (2015). *RepeatMasker Open-4.0* [Computer software]. repeatmasker.org
- Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, *23*(1), 23–35.

- Smith, K. D., Young, K. E., Talbot, C. C., & Schmeckpeper, B. J. (1987). Repeated DNA of the human Y chromosome. *Development*, *101*(Supplement), 77–92. <https://doi.org/10.1242/dev.101.Supplement.77>
- Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., & Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature*, *321*(6071), 674–679. <https://doi.org/10.1038/321674a0>
- Smith, S., Sandoval-Castellanos, E., Lagerholm, V. K., Napierala, H., Sablin, M., Von Seth, J., Fladerer, F. A., Germonpré, M., Wojtal, P., Miller, R., Stewart, J. R., & Dalén, L. (2017). Nonreceding hare lines: Genetic continuity since the Late Pleistocene in European mountain hares (*Lepus timidus*). *Biological Journal of the Linnean Society*, *120*(4), 891–908. <https://doi.org/10.1093/biolinnean/blw009>
- Smyser, T. J., Tabak, M. A., Slotmaker, C., Robeson, M. S., Miller, R. S., Bosse, M., Megens, H.-J., Groenen, M. A. M., Paiva, S. R., de Faria, D. A., Blackburn, H. D., Schmit, B. S., & Piaggio, A. J. (2020). Mixed ancestry from wild and domestic lineages contributes to the rapid expansion of invasive feral swine. *Molecular Ecology*, *29*(6), 1103–1119. <https://doi.org/10.1111/mec.15392>
- 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>
- Souto, J., Marques, J. P., Farelo, L., Costa, J., Queirós, J., Pietri, C., Ballesteros, F., Alves, P. C., Boursot, P., & Melo-Ferreira, J. (2025). Multiple genomic ancestries in the broom hare mark the complex biogeographic history of hares in the Iberian Peninsula. *Evolution*, qpaf082. <https://doi.org/10.1093/evolut/qpaf082>
- Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T. I., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M., & Lancet, D. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current Protocols in Bioinformatics*, *54*(1). <https://doi.org/10.1002/cpbi.5>
- Suchard, M. A., Weiss, R. E., Dorman, K. S., & Sinsheimer, J. S. (2002). Oh Brother, Where Art Thou? A Bayes Factor Test for Recombination with Uncertain Heritage. *Systematic Biology*, *51*(5), 715–728. <https://doi.org/10.1080/10635150290102384>
- Symington, L. S., & Gautier, J. (2011). Double-Strand Break End Resection and Repair Pathway Choice. *Annual Review of Genetics*, *45*(1), 247–271. <https://doi.org/10.1146/annurev-genet-110410-132435>
- Tang, H., Coram, M., Wang, P., Zhu, X., & Risch, N. (2006). Reconstructing Genetic Ancestry Blocks in Admixed Individuals. *The American Journal of Human Genetics*, *79*(1), 1–12. <https://doi.org/10.1086/504302>
- 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>
- Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution*, *3*(2), 170–177. <https://doi.org/10.1038/s41559-018-0777-y>

- Thorvaldsdottir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192. <https://doi.org/10.1093/bib/bbs017>
- Thresher, R., & Griffith, J. (1992). [24] Electron microscopic visualization of DNA and DNA-protein complexes as adjunct to biochemical studies. In *Methods in Enzymology* (Vol. 211, pp. 481–490). Elsevier. [https://doi.org/10.1016/0076-6879\(92\)11026-F](https://doi.org/10.1016/0076-6879(92)11026-F)
- Thulin, C. -G., Jaarola, M., & Tegelström, H. (1997). The occurrence of mountain hare mitochondrial DNA in wild brown hares. *Molecular Ecology*, *6*(5), 463–467. <https://doi.org/10.1046/j.1365-294X.1997.t01-1-00199.x>
- Thulin, C., & Tegelström, H. (2002). Biased geographical distribution of mitochondrial DNA that passed the species barrier from mountain hares to brown hares (genus *Lepus*): An effect of genetic incompatibility and mating behaviour? *Journal of Zoology*, *258*(3), 299–306. <https://doi.org/10.1017/S0952836902001425>
- Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., Heredia, S. M., Hahn, M. A., Caseys, C., Bock, D. G., & Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, *9*(7), 892–908. <https://doi.org/10.1111/eva.12367>
- Torregrosa-Muñumer, R., Goffart, S., Haikonen, J. A., & Pohjoismäki, J. L. O. (2015). Low doses of ultraviolet radiation and oxidative damage induce dramatic accumulation of mitochondrial DNA replication intermediates, fork regression, and replication initiation shift. *Molecular Biology of the Cell*, *26*(23), 4197–4208. <https://doi.org/10.1091/mbc.e15-06-0390>
- Torregrosa-Muñumer, R., Hangas, A., Goffart, S., Blei, D., Zsurka, G., Griffith, J., Kunz, W. S., & Pohjoismäki, J. L. O. (2019). Replication fork rescue in mammalian mitochondria. *Scientific Reports*, *9*(1), 8785. <https://doi.org/10.1038/s41598-019-45244-6>
- Ungar, R. A., Goddard, P. C., Jensen, T. D., Degalez, F., Smith, K. S., Jin, C. A., Undiagnosed Diseases Network, Bonner, D. E., Bernstein, J. A., Wheeler, M. T., & Montgomery, S. B. (2024). Impact of genome build on RNA-seq interpretation and diagnostics. *medRxiv: The Preprint Server for Health Sciences*, 2024.01.11.24301165. <https://doi.org/10.1101/2024.01.11.24301165>
- Valiente-Mullor, C., Beamud, B., Ansari, I., Francés-Cuesta, C., García-González, N., Mejía, L., Ruiz-Hueso, P., & González-Candelas, F. (2021). One is not enough: On the effects of reference genome for the mapping and subsequent analyses of short-reads. *PLoS Computational Biology*, *17*(1), e1008678. <https://doi.org/10.1371/journal.pcbi.1008678>
- Van Bemmelen, J., Nika, I., & Baaijens, J. A. (2025). *Benchmarking the impact of reference genome selection on taxonomic profiling accuracy*. *Bioinformatics*. <https://doi.org/10.1101/2025.02.07.637076>
- 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>
- Vasimuddin, Md., Misra, S., Li, H., & Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 314–324. <https://doi.org/10.1109/IPDPS.2019.00041>
- Verardi, A., Lucchini, V., & Randi, E. (2006). Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (*Canis lupus*) by admixture linkage disequilibrium

- analysis. *Molecular Ecology*, 15(10), 2845–2855. <https://doi.org/10.1111/j.1365-294X.2006.02995.x>
- 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)
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, 33(14), 2202–2204. <https://doi.org/10.1093/bioinformatics/btx153>
- Wallace, J. G., & Mitchell, S. E. (2017). Genotyping-by-Sequencing. *Current Protocols in Plant Biology*, 2(1), 64–77. <https://doi.org/10.1002/cppb.20042>
- Wang, D., Xie, K., Wang, Y., Hu, J., Li, W., Yang, A., Zhang, Q., Ning, C., & Fan, X. (2022). Cost-effectively dissecting the genetic architecture of complex wool traits in rabbits by low-coverage sequencing. *Genetics, Selection, Evolution: GSE*, 54(1), 75. <https://doi.org/10.1186/s12711-022-00766-y>
- Weir, B. S., & Cockerham, C. C. (1984). ESTIMATING F -STATISTICS FOR THE ANALYSIS OF POPULATION STRUCTURE. *Evolution*, 38(6), 1358–1370. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>
- Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P.-C., Hall, R. J., Concepcion, G. T., Ebler, J., Functammasan, A., Kolesnikov, A., Olson, N. D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.-S., Phillippy, A. M., Schatz, M. C., Myers, G., DePristo, M. A., ... Hunkapiller, M. W. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology*, 37(10), 1155–1162. <https://doi.org/10.1038/s41587-019-0217-9>
- Wenger, S. L., Senft, J. R., Sargent, L. M., Bamezai, R., Bairwa, N., & Grant, S. G. (2004). Comparison of Established Cell Lines at Different Passages by Karyotype and Comparative Genomic Hybridization. *Bioscience Reports*, 24(6), 631–639. <https://doi.org/10.1007/s10540-005-2797-5>
- Whibley, A., Kelley, J. L., & Narum, S. R. (2021). The changing face of genome assemblies: Guidance on achieving high-quality reference genomes. *Molecular Ecology Resources*, 21(3), 641–652. <https://doi.org/10.1111/1755-0998.13312>
- Wilson, B. A., Petrov, D. A., & Messer, P. W. (2014). Soft Selective Sweeps in Complex Demographic Scenarios. *Genetics*, 198(2), 669–684. <https://doi.org/10.1534/genetics.114.165571>
- Wiuf, C., Zhao, K., Innan, H., & Nordborg, M. (2004). The probability and chromosomal extent of trans-specific polymorphism. *Genetics*, 168(4), 2363–2372. <https://doi.org/10.1534/genetics.104.029488>
- Wu, L., & Hickson, I. D. (2003). The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature*, 426(6968), 870–874. <https://doi.org/10.1038/nature02253>
- Wu, M. Y., Low, G. W., Forcina, G., van Grouw, H., Lee, B. P. Y.-H., Oh, R. R. Y., & Rheindt, F. E. (2020). Historic and modern genomes unveil a domestic introgression gradient in a wild red junglefowl population. *Evolutionary Applications*, 13(9), 2300–2315. <https://doi.org/10.1111/eva.13023>

- Wu, Y., Xia, L., Zhang, Q., Yang, Q., & Meng, X. (2011). Bidirectional introgressive hybridization between *Lepus capensis* and *Lepus yarkandensis*. *Molecular Phylogenetics and Evolution*, *59*(3), 545–555. <https://doi.org/10.1016/j.ympev.2011.03.027>
- Xia, X., Zhang, S., Zhang, H., Zhang, Z., Chen, N., Li, Z., Sun, H., Liu, X., Lyu, S., Wang, X., Li, Z., Yang, P., Xu, J., Ding, X., Shi, Q., Wang, E., Ru, B., Xu, Z., Lei, C., ... Huang, Y. (2021). Assessing genomic diversity and signatures of selection in Jiaxian Red cattle using whole-genome sequencing data. *BMC Genomics*, *22*(1), 43. <https://doi.org/10.1186/s12864-020-07340-0>
- Yao, X., Tan, Y., Yang, J., Wang, Y., Corlett, R. T., & Manen, J.-F. (2019). Exceptionally high rates of positive selection on the *rbcL* gene in the genus *Ilex* (Aquifoliaceae). *BMC Evolutionary Biology*, *19*(1), 192. <https://doi.org/10.1186/s12862-019-1521-1>
- Zardoni, L., Nardini, E., & Liberi, G. (2020). 2D Gel Electrophoresis to Detect DNA Replication and Recombination Intermediates in Budding Yeast. In K. Hanada (Ed.), *DNA Electrophoresis* (Vol. 2119, pp. 43–59). Springer US. https://doi.org/10.1007/978-1-0716-0323-9_4
- Zbinden, Z. D., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2023). A community genomics approach to natural hybridization. *Proceedings of the Royal Society B: Biological Sciences*, *290*(1999), 20230768. <https://doi.org/10.1098/rspb.2023.0768>
- Zhang, H., Li, X., & Lin, L. (2025). Biological Functions and Clinical Implications of CFLAR: From Cell Death Mechanisms to Therapeutic Targeting in Immune Regulation. *Journal of Inflammation Research, Volume 18*, 4911–4928. <https://doi.org/10.2147/JIR.S519885>
- Zhang, JianZhong., Fang, Yu., Hou, J. Y., Ren, H. Ji., Jiang, Rong., Roos, Pieter., & Dovichi, N. J. (1995). Use of Non-Cross-Linked Polyacrylamide for Four-Color DNA Sequencing by Capillary Electrophoresis Separation of Fragments up to 640 Bases in Length in Two Hours. *Analytical Chemistry*, *67*(24), 4589–4593. <https://doi.org/10.1021/ac00120a026>
- Zhao, L. (2019). Mitochondrial DNA degradation: A quality control measure for mitochondrial genome maintenance and stress response. In *The Enzymes* (Vol. 45, pp. 311–341). Elsevier. <https://doi.org/10.1016/bs.enz.2019.08.004>
- Zheng, Z., Wang, X., Li, M., Li, Y., Yang, Z., Wang, X., Pan, X., Gong, M., Zhang, Y., Guo, Y., Wang, Y., Liu, J., Cai, Y., Chen, Q., Okpeku, M., Colli, L., Cai, D., Wang, K., Huang, S., ... Jiang, Y. (2020). The origin of domestication genes in goats. *Science Advances*, *6*(21), eaaz5216. <https://doi.org/10.1126/sciadv.aaz5216>
- 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>
- Zimin, A. V., Puiu, D., Luo, M.-C., Zhu, T., Koren, S., Marçais, G., Yorke, J. A., Dvořák, J., & Salzberg, S. L. (2017). Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Research*, *27*(5), 787–792. <https://doi.org/10.1101/gr.213405.116>
- 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., Kraysberg, 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>

M2. ORYCTOLAGUS CUNICULUS MINTÁK JEGYZÉKE

SRA Project ID	SRA SampleID	Triviális név	Csoport	Publikáció azonosító	Megjegyzés
PRJNA354575	SRS1819859	Nldwarf_1	NLdwarf	PMID: 27986804	
PRJNA354575	SRS1819864	Nldwarf_2	NLdwarf	PMID: 27986804	
PRJNA354575	SRS1819877	Nldwarf_3	NLdwarf	PMID: 27986804	
PRJNA242290	SRS609654	wildIB_0	wildIB	PMID: 27986804	
PRJNA242290	SRS609655	FrenchLop	FrenchLop	PMID: 25170157	
PRJNA242290	SRS609658	wildIB_1	wildIB	PMID: 25170157	
PRJNA242290	SRS609659	wildIB_2	wildIB	PMID: 25170157	
PRJNA242290	SRS609660	wildIB_3	wildIB	PMID: 25170157	
PRJNA242290	SRS609661	wildIB_4	wildIB	PMID: 25170157	
PRJNA242290	SRS609662	Dutch	Dutch	PMID: 25170157	
PRJNA242290	SRS609663	wildIB_5	wildIB	PMID: 25170157	
PRJNA242290	SRS609664	wildIB_6	wildIB	PMID: 25170157	
PRJNA242290	SRS609665	wildFR_1	wildFR	PMID: 25170157	
PRJNA242290	SRS609666	wildIB_7	wildIB	PMID: 25170157	
PRJNA242290	SRS609668	Champagne	Champagne	PMID: 25170157	
PRJNA242290	SRS609669	wildFR_2	wildFR	PMID: 25170157	
PRJNA242290	SRS609670	wildIB_8	wildIB	PMID: 25170157	
PRJNA242290	SRS609671	wildFR_3	wildFR	PMID: 25170157	
PRJNA242290	SRS609672	NZW_pool	NewZealandWhite_public	PMID: 25170157	
PRJNA242290	SRS609673	FlemishGiant	FlemishGiant	PMID: 25170157	
PRJNA242290	SRS609674	wildIB_9	WildIB	PMID: 25170157	
PRJNA242290	SRS609675	BelgianHare	BelgianHare	PMID: 25170157	
PRJNA242290	SRS609676	wildIB_10	wildIB	PMID: 25170157	
PRJNA242290	SRS609677	wildIB_11	wildIB	PMID: 25170157	
PRJNA274594	SRS837260	japanesewhite_1	JapaneseWhite	PMID: 27245873	
PRJNA274594	SRS837261	japanesewhite_2	JapaneseWhite	PMID: 27245873	
PRJNA274594	SRS837262	japanesewhite_3	JapaneseWhite	PMID: 27245873	
PRJNA274594	SRS837263	japanesewhite_4	JapaneseWhite	PMID: 27245873	
PRJNA274594	SRS837264	japanesewhite_5	JapaneseWhite	PMID: 27245873	
PRJNA274594	SRS837265	japanesewhite_6	JapaneseWhite	PMID: 27245873	
PRJNA274594	SRS837266	japanesewhite_7	JapaneseWhite	PMID: 27245873	
PRJNA274594	SRS837267	japanesewhite_8	JapaneseWhite	PMID: 27245873	
PRJNA274594	SRS837268	japanesewhite_9	JapaneseWhite	PMID: 27245873	
PRJNA274594	SRS837271	Watanabe_2	Watanabe	PMID: 27245873	

PRJNA274594	SRS837272	Watanabe_3	Watanabe	PMID: 27245873	
PRJNA274594	SRS837273	Watanabe_4	Watanabe	PMID: 27245873	
PRJNA274594	SRS837274	Watanabe_5	Watanabe	PMID: 27245873	
PRJNA274594	SRS837275	Watanabe_6	Watanabe	PMID: 27245873	
PRJNA274594	SRS837276	Watanabe_7	Watanabe	PMID: 27245873	
PRJNA274594	SRS837277	Watanabe_8	Watanabe	PMID: 27245873	
PRJNA274594	SRS837278	Watanabe_9	Watanabe	PMID: 27245873	
PRJNA274594	SRS837279	Watanabe_10	Watanabe	PMID: 27245873	
PRJNA274594	SRS837280	NZW_p1	NewZealandWhite_public	PMID: 27245873	
PRJNA274594	SRS837281	NZW_p2	NewZealandWhite_public	PMID: 27245873	
PRJNA274594	SRS837282	NZW_p3	NewZealandWhite_public	PMID: 27245873	
PRJNA274594	SRS837283	NZW_p4	NewZealandWhite_public	PMID: 27245873	
PRJNA274594	SRS837284	NZW_p5	NewZealandWhite_public	PMID: 27245873	
PRJNA274594	SRS837285	NZW_p6	NewZealandWhite_public	PMID: 27245873	
PRJNA274594	SRS837286	NZW_p7	NewZealandWhite_public	PMID: 27245873	
PRJNA274594	SRS837287	NZW_p8	NewZealandWhite_public	PMID: 27245873	
PRJNA274594	SRS837288	NZW_p9	NewZealandWhite_public	PMID: 27245873	
PRJNA274594	SRS837289	NZW_p10	NewZealandWhite_public	PMID: 27245873	
PRJNA274594	SRS837290	NZW_p11	NewZealandWhite_public	PMID: 27245873	
PRJEB28783	ERS2757785	Satinangora_1	Angora	Ismeretlen	
PRJEB28783	ERS2757786	Satinangora_2	Angora	Ismeretlen	
PRJNA1118464	SAMN41610687	wild_1	WildHUa	PMID: 40150307	Általunk szekvenált
PRJNA1118464	SAMN41610688	wild_2	WildHUa	PMID: 40150308	Általunk szekvenált
PRJNA1118464	SAMN41610689	wild_3	WildHUa	PMID: 40150309	Általunk szekvenált
PRJNA1118464	SAMN41610690	wild_4	WildHUa	PMID: 40150310	Általunk szekvenált
PRJNA1118464	SAMN41610691	wild_5	WildHUa	PMID: 40150311	Általunk szekvenált
PRJNA1118464	SAMN41610692	wild_6	WildHUa	PMID: 40150312	Általunk szekvenált
PRJNA1118464	SAMN41610693	Zika_1	Zika	PMID: 40150313	Általunk szekvenált
PRJNA1118464	SAMN41610694	Zika_2	Zika	PMID: 40150314	Általunk szekvenált

PRJNA1118464	SAMN4161069 5	Zika_3	Zika	PMID: 40150315	Átalunk szekvenált
PRJNA1118464	SAMN4161069 6	Zika_4	Zika	PMID: 40150316	Átalunk szekvenált
PRJNA1118464	SAMN4161069 7	Zika_5	Zika	PMID: 40150317	Átalunk szekvenált
PRJNA1118464	SAMN4161069 8	Zika_6	Zika	PMID: 40150318	Átalunk szekvenált
PRJNA1118464	SAMN4161069 9	Rex_1	Rex	PMID: 40150319	Átalunk szekvenált
PRJNA1118464	SAMN4161070 0	Rex_2	Rex	PMID: 40150320	Átalunk szekvenált
PRJNA1118464	SAMN4161070 1	Rex_3	Rex	PMID: 40150321	Átalunk szekvenált
PRJNA1118464	SAMN4161070 2	Rex_4	Rex	PMID: 40150322	Átalunk szekvenált
PRJNA1118464	SAMN4161070 3	Rex_5	Rex	PMID: 40150323	Átalunk szekvenált
PRJNA1118464	SAMN4161070 4	Californian_1	Californian	PMID: 40150324	Átalunk szekvenált
PRJNA1118464	SAMN4161070 5	Californian_2	Californian	PMID: 40150325	Átalunk szekvenált
PRJNA1118464	SAMN4161070 6	Californian_3	Californian	PMID: 40150326	Átalunk szekvenált
PRJNA1118464	SAMN4161070 7	NewZealandWhite_1	NZW	PMID: 40150327	Átalunk szekvenált
PRJNA1118464	SAMN4161070 8	NewZealandWhite_2	NZW	PMID: 40150328	Átalunk szekvenált
PRJNA1118464	SAMN4161070 9	NewZealandWhite_3	NZW	PMID: 40150329	Átalunk szekvenált
PRJNA1118464	SAMN4161071 0	Thuringian	Thuringian	PMID: 40150330	Átalunk szekvenált
PRJNA1118464	SAMN4161071 1	wild_7	wildHUb	PMID: 40150331	Átalunk szekvenált
PRJNA1118464	SAMN4161071 2	wild_8	wildHUb	PMID: 40150332	Átalunk szekvenált
PRJNA1118464	SAMN4161071 3	wild_9	wildHUb	PMID: 40150333	Átalunk szekvenált
PRJNA1118464	SAMN4161071 4	wild_10	wildHUb	PMID: 40150334	Átalunk szekvenált
PRJNA1118464	SAMN4161071 5	wild_11	wildHUb	PMID: 40150335	Átalunk szekvenált
PRJNA1118464	SAMN4161071 6	wild_12	wildHUb	PMID: 40150336	Átalunk szekvenált
PRJNA1118464	SAMN4161071 7	wild_13	wildHUb	PMID: 40150337	Átalunk szekvenált
PRJNA1118464	SAMN4161071 8	wild_14	wildHUb	PMID: 40150338	Átalunk szekvenált
PRJNA1118464	SAMN4161071 9	Hycole_1	Hycole	PMID: 40150339	Átalunk szekvenált
PRJNA1118464	SAMN4161072 0	Hycole_2	Hycole	PMID: 40150340	Átalunk szekvenált
PRJNA1118464	SAMN4161072 1	Hycole_3	Hycole	PMID: 40150341	Átalunk szekvenált
PRJNA1118464	SAMN4161072 2	Hycole_4	Hycole	PMID: 40150342	Átalunk szekvenált
PRJNA1118464	SAMN4161072 3	Hycole_5	Hycole	PMID: 40150343	Átalunk szekvenált
PRJNA1118464	SAMN4161072 4	Hycole_6	Hycole	PMID: 40150344	Átalunk szekvenált

PRJNA1118464	SAMN4161072 5	Hycole 7	Hycole	PMID: 40150345	Átalunk szekvenált
PRJNA1118464	SAMN4161072 6	Hycole 8	Hycole	PMID: 40150346	Átalunk szekvenált
PRJNA1118464	SAMN4161072 7	Hycole 9	Hycole	PMID: 40150347	Átalunk szekvenált
PRJNA1118464	SAMN4161072 8	Hycole 10	Hycole	PMID: 40150348	Átalunk szekvenált

M3. ERŐSEN DIFFERENCIÁLT VARIÁCIÓK A SZELEKTÁLT RÉGIÓKBAN

Genomi pozíció		Allél		Allélfrekvencia				Fst
Kromoszóma	Pozíció	Referencia	Alternatív	Házi_allél1	Házi_allél2	Üregi_allél1	Üregi_allél2	
13	128068123	T	C	0	1	1	0	1
13	128062970	G	A	0	1	1	0	1
13	128065406	G	A	0.00735294	0.992647	1	0	0.988805
7	141926050	A	G	1	0	0.0227273	0.977273	0.988633
7	141914463	C	G	1	0	0.0227273	0.977273	0.988633
7	141910838	T	C	1	0	0.0227273	0.977273	0.988633
7	141891367	T	C	1	0	0.0227273	0.977273	0.988633
7	141891201	A	G	1	0	0.0227273	0.977273	0.988633
7	141884883	T	C	1	0	0.0227273	0.977273	0.988633
7	141883795	T	C	1	0	0.0227273	0.977273	0.988633
14	77819024	A	C	1	0	0.0227273	0.977273	0.988633
13	139087024	G	A	0	1	0.977273	0.0227273	0.988633
13	139066102	T	C	1	0	0.0227273	0.977273	0.988633
13	128056470	C	T	0	1	0.977273	0.0227273	0.988633
13	128065376	C	T	0.0147059	0.985294	1	0	0.977695
13	139101914	G	A	0.00735294	0.992647	0.977273	0.0227273	0.977006
13	139084246	G	A	0.00735294	0.992647	0.977273	0.0227273	0.977006
13	139071889	C	G	0.00735294	0.992647	0.977273	0.0227273	0.977006
13	139050301	G	C	0.00735294	0.992647	0.977273	0.0227273	0.977006
13	139043297	C	T	0.00735294	0.992647	0.977273	0.0227273	0.977006
13	128061353	C	T	0.00735294	0.992647	0.977273	0.0227273	0.977006
7	141925216	G	T	1	0	0.0454545	0.954545	0.977001
7	141920594	T	C	1	0	0.0454545	0.954545	0.977001
7	141919718	T	C	1	0	0.0454545	0.954545	0.977001
7	141918730	C	T	1	0	0.0454545	0.954545	0.977001
7	141918727	C	T	1	0	0.0454545	0.954545	0.977001
7	141918609	A	T	1	0	0.0454545	0.954545	0.977001
7	141918455	C	T	1	0	0.0454545	0.954545	0.977001
7	141910973	T	C	1	0	0.0454545	0.954545	0.977001
7	141908434	C	CA	1	0	0.0454545	0.954545	0.977001
7	141889754	T	C	1	0	0.0454545	0.954545	0.977001
14	77796671	C	T	1	0	0.0454545	0.954545	0.977001
14	77792627	T	C	1	0	0.0454545	0.954545	0.977001
14	77787141	A	G	1	0	0.0454545	0.954545	0.977001
14	77764296	C	T	1	0	0.0454545	0.954545	0.977001
13	128055790	A	AT	0	1	0.954545	0.0454545	0.977001
12	83342442	C	T	1	0	0.0454545	0.954545	0.977001
7	141928602	A	G	0.985294	0.0147059	0.0227273	0.977273	0.96547
13	139098828	T	C	0.0147059	0.985294	0.977273	0.0227273	0.96547
13	139091173	G	T	0.0147059	0.985294	0.977273	0.0227273	0.96547
13	139085609	A	G	0.0147059	0.985294	0.977273	0.0227273	0.96547
13	139081708	A	G	0.0147059	0.985294	0.977273	0.0227273	0.96547
13	139079452	T	C	0.0147059	0.985294	0.977273	0.0227273	0.96547
13	139038418	C	T	0.0147059	0.985294	0.977273	0.0227273	0.96547

7	141920984	T	A	1	0	0.0681818	0.931818	0.965095
7	141919144	G	A	1	0	0.0681818	0.931818	0.965095
7	141902737	GC	G	1	0	0.0681818	0.931818	0.965095
7	141897372	G	A	1	0	0.0681818	0.931818	0.965095
7	141896627	C	CT	1	0	0.0681818	0.931818	0.965095
14	77811899	CT	C	1	0	0.0681818	0.931818	0.965095
14	77810142	A	G	1	0	0.0681818	0.931818	0.965095
14	77805815	A	G	1	0	0.0681818	0.931818	0.965095
14	77802591	T	C	1	0	0.0681818	0.931818	0.965095
14	77786088	A	C	1	0	0.0681818	0.931818	0.965095
14	29962506	G	A	1	0	0.0681818	0.931818	0.965095
14	29956941	A	ACTG	1	0	0.0681818	0.931818	0.965095
14	2448350	T	C	1	0	0.0681818	0.931818	0.965095
13	139038836	C	T	0	1	0.931818	0.0681818	0.965095
13	139038834	C	CT	0	1	0.931818	0.0681818	0.965095
13	139074191	A	G	0.00735294	0.992647	0.954545	0.0454545	0.964924
13	139064536	C	T	0.00735294	0.992647	0.954545	0.0454545	0.964924
13	139057794	C	T	0.00735294	0.992647	0.954545	0.0454545	0.964924
13	128590248	G	C	0.992647	0.00735294	0.0454545	0.954545	0.964924
13	128064227	C	T	0.00735294	0.992647	0.954545	0.0454545	0.964924
13	128060204	CGG	C	0.00735294	0.992647	0.954545	0.0454545	0.964924
13	128056036	T	G	0.00735294	0.992647	0.954545	0.0454545	0.964924
7	141881340	G	A	0.977941	0.0220588	0.0227273	0.977273	0.954025
13	139081516	T	C	0.0220588	0.977941	0.977273	0.0227273	0.954025
13	139065522	C	T	0.0147059	0.985294	0.954545	0.0454545	0.952947
13	139050311	C	T	0.0147059	0.985294	0.954545	0.0454545	0.952947
13	139046812	C	T	0.0147059	0.985294	0.954545	0.0454545	0.952947
13	139043175	TCTC	T	0.0147059	0.985294	0.954545	0.0454545	0.952947
12	83347145	T	C	0.985294	0.0147059	0.0454545	0.954545	0.952947
7	141922878	T	C	1	0	0.0909091	0.909091	0.952905
7	141913075	C	T	1	0	0.0909091	0.909091	0.952905
7	141906969	A	G	1	0	0.0909091	0.909091	0.952905
7	141902746	T	G	1	0	0.0909091	0.909091	0.952905
7	141902612	CT	C	1	0	0.0909091	0.909091	0.952905
7	141895675	T	C	1	0	0.0909091	0.909091	0.952905
7	141892696	C	A	1	0	0.0909091	0.909091	0.952905
20	6322474	A	C	1	0	0.0909091	0.909091	0.952905
20	6322438	A	G	1	0	0.0909091	0.909091	0.952905
14	77814761	A	G	1	0	0.0909091	0.909091	0.952905
14	77803046	T	C	1	0	0.0909091	0.909091	0.952905
14	77795313	G	A	1	0	0.0909091	0.909091	0.952905
14	77785635	A	G	1	0	0.0909091	0.909091	0.952905
14	77785368	T	C	1	0	0.0909091	0.909091	0.952905
14	77757294	G	GA	1	0	0.0909091	0.909091	0.952905
14	29962875	A	G	1	0	0.0909091	0.909091	0.952905
14	2450139	A	C	1	0	0.0909091	0.909091	0.952905
12	83385829	G	T	1	0	0.0909091	0.909091	0.952905
12	83350636	T	C	1	0	0.0909091	0.909091	0.952905
12	83344394	T	G	1	0	0.0909091	0.909091	0.952905

14	77815476	G	A	1	0	0.0909091	0.909091	0.952609
14	77792915	C	T	0.992647	0.00735294	0.0681818	0.931818	0.952551
14	2449012	A	G	0.992647	0.00735294	0.0681818	0.931818	0.952551
14	2448845	C	T	0.992647	0.00735294	0.0681818	0.931818	0.952551
13	139052567	C	T	0.00735294	0.992647	0.931818	0.0681818	0.952551
13	139039790	C	T	0.00735294	0.992647	0.931818	0.0681818	0.952551
13	139039789	C	A	0.00735294	0.992647	0.931818	0.0681818	0.952551
13	139039704	C	T	0.00735294	0.992647	0.931818	0.0681818	0.952551
12	83310444	T	C	0.992647	0.00735294	0.0681818	0.931818	0.952551
7	141883652	A	T	0.970588	0.0294118	0.0227273	0.977273	0.942669
13	139091333	C	T	0.0294118	0.970588	0.977273	0.0227273	0.942669
13	139096668	CTGCCTGTAA	C	0.0220588	0.977941	0.954545	0.0454545	0.941068
13	128592119	A	G	0.977941	0.0220588	0.0454545	0.954545	0.941068
7	141913325	T	C	1	0	0.113636	0.886364	0.940422
7	141892338	A	G	1	0	0.113636	0.886364	0.940422
20	6293332	G	C	1	0	0.113636	0.886364	0.940422
14	77805068	A	G	1	0	0.113636	0.886364	0.940422
14	77786089	T	C	1	0	0.113636	0.886364	0.940422
14	77784808	C	T	1	0	0.113636	0.886364	0.940422
14	77760484	A	G	1	0	0.113636	0.886364	0.940422
14	29962805	A	G	1	0	0.113636	0.886364	0.940422
14	29962558	T	A	1	0	0.113636	0.886364	0.940422
14	29962222	C	T	1	0	0.113636	0.886364	0.940422
14	29958419	T	C	1	0	0.113636	0.886364	0.940422
14	29947697	T	C	1	0	0.113636	0.886364	0.940422
14	2448342	G	T	1	0	0.113636	0.886364	0.940422
12	83378072	G	A	1	0	0.113636	0.886364	0.940422
12	83378009	G	T	1	0	0.113636	0.886364	0.940422
12	83377341	G	A	1	0	0.113636	0.886364	0.940422
13	139072706	C	T	0.0147059	0.985294	0.931818	0.0681818	0.940115
13	128593936	T	C	0.985294	0.0147059	0.0681818	0.931818	0.940115
13	128593934	T	C	0.985294	0.0147059	0.0681818	0.931818	0.940115
13	128057957	A	G	0.0147059	0.985294	0.931818	0.0681818	0.940115
7	141925058	T	A	1	0	0.113636	0.886364	0.940113
7	141918250	G	C	1	0	0.113636	0.886364	0.940113
14	2448509	T	C	0.992647	0.00735294	0.0909091	0.909091	0.939874
14	2447912	A	C	0.992647	0.00735294	0.0909091	0.909091	0.939874
13	139063797	C	T	0.00735294	0.992647	0.909091	0.0909091	0.939874
13	139047717	G	C	0.00735294	0.992647	0.909091	0.0909091	0.939874
13	139041759	ATGTC	A	0.00735294	0.992647	0.909091	0.0909091	0.939874
13	139039728	C	T	0.00735294	0.992647	0.909091	0.0909091	0.939874
13	139038354	A	G	0.00735294	0.992647	0.909091	0.0909091	0.939874
13	128064178	A	G	0.00735294	0.992647	0.909091	0.0909091	0.939874
13	128055596	T	C	0.00735294	0.992647	0.909091	0.0909091	0.939874
7	141868397	TTTTTA	T	0.970588	0.0294118	0.0454545	0.954545	0.929287
7	141913182	A	G	1	0	0.136364	0.863636	0.927635
7	141913102	A	T	1	0	0.136364	0.863636	0.927635
7	141903741	T	C	1	0	0.136364	0.863636	0.927635
7	141902771	A	G	1	0	0.136364	0.863636	0.927635

7	141897807	A	G	1	0	0.136364	0.863636	0.927635
7	141897357	G	T	1	0	0.136364	0.863636	0.927635
7	141883933	A	C	1	0	0.136364	0.863636	0.927635
20	6323746	A	G	1	0	0.136364	0.863636	0.927635
20	6323201	A	G	1	0	0.136364	0.863636	0.927635
20	6322100	A	G	1	0	0.136364	0.863636	0.927635
20	6321859	A	C	1	0	0.136364	0.863636	0.927635
20	6310778	C	T	1	0	0.136364	0.863636	0.927635
14	77804341	T	C	1	0	0.136364	0.863636	0.927635
14	77795791	A	C	1	0	0.136364	0.863636	0.927635
14	77782138	T	C	1	0	0.136364	0.863636	0.927635
14	77760450	T	A	1	0	0.136364	0.863636	0.927635
14	29957962	A	G	1	0	0.136364	0.863636	0.927635
14	29944537	A	G	1	0	0.136364	0.863636	0.927635
12	83378654	G	A	1	0	0.136364	0.863636	0.927635
12	83377275	A	G	1	0	0.136364	0.863636	0.927635
12	83366207	G	C	1	0	0.136364	0.863636	0.927635
12	83365851	A	G	1	0	0.136364	0.863636	0.927635
12	83365621	T	C	1	0	0.136364	0.863636	0.927635
12	83353269	A	G	1	0	0.136364	0.863636	0.927635
7	141894110	G	A	1	0	0.136364	0.863636	0.927313
14	29950742	T	C	1	0	0.136364	0.863636	0.927313
14	29947598	G	A	1	0	0.136364	0.863636	0.927313
4	68737358	T	C	0.985294	0.0147059	0.0909091	0.909091	0.926962
4	68494032	C	T	0.985294	0.0147059	0.0909091	0.909091	0.926962
13	139091177	A	G	0.0147059	0.985294	0.0909091	0.0909091	0.926962
13	128593888	A	G	0.985294	0.0147059	0.0909091	0.909091	0.926962
13	128057967	G	T	0.0147059	0.985294	0.0909091	0.0909091	0.926962
13	128057380	C	T	0.0147059	0.985294	0.0909091	0.0909091	0.926962
13	128055377	A	G	0.0147059	0.985294	0.0909091	0.0909091	0.926962
4	68496512	A	G	0.992647	0.00735294	0.113636	0.886364	0.926885
12	83317307	G	A	0.992647	0.00735294	0.113636	0.886364	0.926885
12	83311861	A	G	0.992647	0.00735294	0.113636	0.886364	0.926885
12	83311509	C	CCA	0.992647	0.00735294	0.113636	0.886364	0.926885
13	139038851	C	T	0.00735294	0.992647	0.886364	0.113636	0.926574
7	141928051	T	C	0.955882	0.0441176	0.0227273	0.977273	0.919948
7	141896791	A	G	0.955882	0.0441176	0.0227273	0.977273	0.919948
7	141866898	T	C	0.963235	0.0367647	0.0454545	0.954545	0.917601
7	141907043	G	A	1	0	0.159091	0.840909	0.914534
7	141892542	A	G	1	0	0.159091	0.840909	0.914534
7	141883932	A	G	1	0	0.159091	0.840909	0.914534
4	68733898	T	C	1	0	0.159091	0.840909	0.914534
20	6313012	A	G	1	0	0.159091	0.840909	0.914534
20	6311329	A	G	1	0	0.159091	0.840909	0.914534
20	6291181	C	T	1	0	0.159091	0.840909	0.914534
14	77821306	C	T	1	0	0.159091	0.840909	0.914534
14	29960826	G	T	1	0	0.159091	0.840909	0.914534
14	29955945	A	G	1	0	0.159091	0.840909	0.914534
14	29947647	A	G	1	0	0.159091	0.840909	0.914534

14	29947611	G	A	1	0	0.159091	0.840909	0.914534
13	139052603	C	A	0	1	0.840909	0.159091	0.914534
12	83379726	C	T	1	0	0.159091	0.840909	0.914534
12	83373607	C	T	1	0	0.159091	0.840909	0.914534
12	83327797	G	T	1	0	0.159091	0.840909	0.914534
12	83327796	T	A	1	0	0.159091	0.840909	0.914534
12	83325469	C	T	1	0	0.159091	0.840909	0.914534
12	83324303	C	G	1	0	0.159091	0.840909	0.914534
12	83319742	C	T	1	0	0.159091	0.840909	0.914534
7	141889209	A	G	1	0	0.159091	0.840909	0.914198
2	97374222	T	C	1	0	0.159091	0.840909	0.914198
14	77780762	C	T	1	0	0.159091	0.840909	0.914198
12	83328081	T	C	1	0	0.159091	0.840909	0.914198
7	134038268	A	G	0.977941	0.0220588	0.0909091	0.909091	0.914168
4	68500359	G	A	0.992647	0.00735294	0.136364	0.863636	0.913571
20	6308323	A	G	0.992647	0.00735294	0.136364	0.863636	0.913571
14	29956651	G	A	0.992647	0.00735294	0.136364	0.863636	0.913571
14	2447324	A	G	0.992647	0.00735294	0.136364	0.863636	0.913571
13	128058026	G	A	0.00735294	0.992647	0.863636	0.136364	0.913571
13	128057566	G	C	0.00735294	0.992647	0.863636	0.136364	0.913571
12	83318224	A	G	0.992647	0.00735294	0.136364	0.863636	0.913571
12	83312614	A	G	0.992647	0.00735294	0.136364	0.863636	0.913571
7	141929397	T	C	0.985294	0.0147059	0.113636	0.886364	0.913478
4	68502918	A	T	0.985294	0.0147059	0.113636	0.886364	0.913478
4	68502711	G	T	0.985294	0.0147059	0.113636	0.886364	0.913478
4	68739807	A	G	0.985294	0.0147059	0.113636	0.886364	0.913165
14	2236100	C	T	0.955882	0.0441176	0.0454545	0.954545	0.906011
7	141927793	T	C	0.970588	0.0294118	0.0909091	0.909091	0.901489
7	141920881	G	A	1	0	0.181818	0.818182	0.901108
7	141913158	C	A	1	0	0.181818	0.818182	0.901108
7	141912348	A	G	1	0	0.181818	0.818182	0.901108
7	141903751	A	G	1	0	0.181818	0.818182	0.901108
7	141900519	T	C	1	0	0.181818	0.818182	0.901108
7	141898083	A	G	1	0	0.181818	0.818182	0.901108
7	141897133	A	G	1	0	0.181818	0.818182	0.901108
20	6322448	A	T	1	0	0.181818	0.818182	0.901108
20	6321678	T	C	1	0	0.181818	0.818182	0.901108
20	6314168	C	CT	1	0	0.181818	0.818182	0.901108
14	77818087	TA	T	1	0	0.181818	0.818182	0.901108
14	77805454	C	T	1	0	0.181818	0.818182	0.901108
14	77805034	T	C	1	0	0.181818	0.818182	0.901108
14	77801118	G	A	1	0	0.181818	0.818182	0.901108
14	77782305	G	A	1	0	0.181818	0.818182	0.901108
14	77782283	A	G	1	0	0.181818	0.818182	0.901108
14	77782273	G	A	1	0	0.181818	0.818182	0.901108
14	77782072	G	A	1	0	0.181818	0.818182	0.901108
14	77774105	G	A	1	0	0.181818	0.818182	0.901108
14	77758973	T	C	1	0	0.181818	0.818182	0.901108
14	77753421	G	A	1	0	0.181818	0.818182	0.901108

14	29961818	G	T	1	0	0.181818	0.818182	0.901108
14	29947829	A	G	1	0	0.181818	0.818182	0.901108
14	29947348	C	CAGGT	1	0	0.181818	0.818182	0.901108
14	29946443	CTA	C	1	0	0.181818	0.818182	0.901108
14	29944652	G	C	1	0	0.181818	0.818182	0.901108
12	83381919	T	C	1	0	0.181818	0.818182	0.901108
12	83378774	T	C	1	0	0.181818	0.818182	0.901108
12	83378017	T	A	1	0	0.181818	0.818182	0.901108
12	83369783	A	G	1	0	0.181818	0.818182	0.901108
12	83323214	C	T	1	0	0.181818	0.818182	0.901108
14	77799815	G	A	1	0	0.181818	0.818182	0.900756
13	128063286	C	T	0	1	0.818182	0.181818	0.900756
18	63112446	A	G	1	0	0.181818	0.818182	0.900403
4	68744818	C	A	0.977941	0.0220588	0.113636	0.886364	0.9002
20	6307100	G	A	0.992647	0.00735294	0.159091	0.840909	0.899921
14	77788293	C	T	0.992647	0.00735294	0.159091	0.840909	0.899921
13	139050044	T	C	0.00735294	0.992647	0.840909	0.159091	0.899921
13	139039784	G	C	0.00735294	0.992647	0.840909	0.159091	0.899921
12	83310155	GA	G	0.992647	0.00735294	0.159091	0.840909	0.899921
4	68505089	A	G	0.985294	0.0147059	0.136364	0.863636	0.89965
4	68500383	AC	A	0.985294	0.0147059	0.136364	0.863636	0.89965
13	128064950	T	TGCGC	0.0147059	0.985294	0.863636	0.136364	0.89965
7	141917862	A	G	0.992647	0.00735294	0.159091	0.840909	0.899581
4	68500661	G	A	0.992647	0.00735294	0.159091	0.840909	0.899581
20	6305709	T	C	0.992647	0.00735294	0.159091	0.840909	0.899581
13	139038408	G	A	0.0147059	0.985294	0.863636	0.136364	0.899323
7	141923359	T	C	1	0	0.204545	0.795455	0.887345
7	141920290	C	T	1	0	0.204545	0.795455	0.887345
7	141911680	T	C	1	0	0.204545	0.795455	0.887345
7	141911529	G	T	1	0	0.204545	0.795455	0.887345
7	141911093	A	T	1	0	0.204545	0.795455	0.887345
20	6321870	A	C	1	0	0.204545	0.795455	0.887345
20	6320167	C	T	1	0	0.204545	0.795455	0.887345
20	6294281	G	T	1	0	0.204545	0.795455	0.887345
20	6292210	T	C	1	0	0.204545	0.795455	0.887345
14	77822858	T	C	1	0	0.204545	0.795455	0.887345
14	77809528	A	G	1	0	0.204545	0.795455	0.887345
14	77808929	T	C	1	0	0.204545	0.795455	0.887345
14	77806686	C	T	1	0	0.204545	0.795455	0.887345
14	77793500	T	C	1	0	0.204545	0.795455	0.887345
14	77785642	A	G	1	0	0.204545	0.795455	0.887345
14	77782024	T	A	1	0	0.204545	0.795455	0.887345
14	29961502	C	A	1	0	0.204545	0.795455	0.887345
14	29958728	T	G	1	0	0.204545	0.795455	0.887345
14	29958701	C	T	1	0	0.204545	0.795455	0.887345
14	29955236	T	C	1	0	0.204545	0.795455	0.887345
14	29949183	C	G	1	0	0.204545	0.795455	0.887345
14	29944496	T	G	1	0	0.204545	0.795455	0.887345
14	29943632	A	G	1	0	0.204545	0.795455	0.887345

14	29943078	A	G	1	0	0.204545	0.795455	0.887345
12	83380892	A	G	1	0	0.204545	0.795455	0.887345
12	83377879	T	C	1	0	0.204545	0.795455	0.887345
12	83376241	C	T	1	0	0.204545	0.795455	0.887345
12	83372817	T	TA	1	0	0.204545	0.795455	0.887345
7	141906859	A	G	1	0	0.204545	0.795455	0.886976
7	141897713	C	T	1	0	0.204545	0.795455	0.886976
4	69076748	A	G	1	0	0.204545	0.795455	0.886976
14	77804041	C	T	1	0	0.204545	0.795455	0.886976
14	77799813	T	C	1	0	0.204545	0.795455	0.886976
14	77786721	A	G	1	0	0.204545	0.795455	0.886976
14	77769962	T	C	1	0	0.204545	0.795455	0.886976
12	83324775	T	G	1	0	0.204545	0.795455	0.886976
14	77805720	C	T	1	0	0.204545	0.795455	0.886607
7	134038202	G	A	0.992647	0.00735294	0.181818	0.818182	0.885922
14	29957016	C	T	0.992647	0.00735294	0.181818	0.818182	0.885922
12	83318367	T	C	0.992647	0.00735294	0.181818	0.818182	0.885922
12	83318318	A	G	0.992647	0.00735294	0.181818	0.818182	0.885922
12	83317423	G	A	0.992647	0.00735294	0.181818	0.818182	0.885922
7	134039441	A	G	0.977941	0.0220588	0.136364	0.863636	0.885871
20	6308493	C	T	0.985294	0.0147059	0.159091	0.840909	0.885465
13	128060344	T	C	0.0147059	0.985294	0.840909	0.159091	0.885124
8	20031725	G	A	0.941176	0.0588235	0.0454545	0.954545	0.882822
8	19620083	C	T	0.941176	0.0588235	0.0454545	0.954545	0.882822
7	141923336	A	G	1	0	0.227273	0.772727	0.873232
7	141923067	T	C	1	0	0.227273	0.772727	0.873232
7	141920407	G	A	1	0	0.227273	0.772727	0.873232
7	141899405	T	C	1	0	0.227273	0.772727	0.873232
7	141898116	C	T	1	0	0.227273	0.772727	0.873232
7	141898060	T	C	1	0	0.227273	0.772727	0.873232
7	141896153	A	C	1	0	0.227273	0.772727	0.873232
20	6318420	A	G	1	0	0.227273	0.772727	0.873232
20	6294285	G	C	1	0	0.227273	0.772727	0.873232
20	6292143	C	T	1	0	0.227273	0.772727	0.873232
20	6291069	G	A	1	0	0.227273	0.772727	0.873232
15	12919198	A	T	1	0	0.227273	0.772727	0.873232
14	77809430	C	T	1	0	0.227273	0.772727	0.873232
14	77809272	G	A	1	0	0.227273	0.772727	0.873232
14	77809163	A	G	1	0	0.227273	0.772727	0.873232
14	77806047	A	G	1	0	0.227273	0.772727	0.873232
14	77795457	T	C	1	0	0.227273	0.772727	0.873232
14	77786130	T	G	1	0	0.227273	0.772727	0.873232
14	29960982	T	C	1	0	0.227273	0.772727	0.873232
14	29958455	T	C	1	0	0.227273	0.772727	0.873232
14	29945774	T	C	1	0	0.227273	0.772727	0.873232
14	29944563	C	A	1	0	0.227273	0.772727	0.873232
12	83382360	G	T	1	0	0.227273	0.772727	0.873232
12	83378069	T	C	1	0	0.227273	0.772727	0.873232
7	141920557	G	A	1	0	0.227273	0.772727	0.872846

7	141905423	T	G	1	0	0.227273	0.772727	0.872846
4	69072923	G	T	1	0	0.227273	0.772727	0.872846
4	69068041	A	T	1	0	0.227273	0.772727	0.872846
20	6313184	G	C	1	0	0.227273	0.772727	0.872846
14	77804588	T	C	1	0	0.227273	0.772727	0.872846
14	77795620	A	G	1	0	0.227273	0.772727	0.872846
14	77793886	T	C	1	0	0.227273	0.772727	0.872846
14	77782129	TC	T	1	0	0.227273	0.772727	0.872846
14	29963442	C	T	1	0	0.227273	0.772727	0.872846
14	29962899	G	A	1	0	0.227273	0.772727	0.872846
14	29961464	C	A	1	0	0.227273	0.772727	0.872846
14	29958702	A	G	1	0	0.227273	0.772727	0.872846
13	128063502	C	T	0	1	0.772727	0.227273	0.872846
12	83324445	GA	G	1	0	0.227273	0.772727	0.872846
4	69064924	A	T	1	0	0.227273	0.772727	0.872459
16	34674174	T	C	1	0	0.227273	0.772727	0.872459
7	141863340	A	C	0.970588	0.0294118	0.136364	0.863636	0.872232
7	134038289	A	G	0.970588	0.0294118	0.136364	0.863636	0.871569
4	68745299	C	CAAT	0.992647	0.00735294	0.204545	0.795455	0.871562
4	68734594	G	A	0.992647	0.00735294	0.204545	0.795455	0.871562
13	139090869	C	T	0.00735294	0.992647	0.795455	0.204545	0.871562
13	139037146	T	C	0.00735294	0.992647	0.795455	0.204545	0.871119
14	29956321	CT	C	0.992647	0.00735294	0.204545	0.795455	0.870819
14	2448446	A	G	0.992647	0.00735294	0.204545	0.795455	0.870819
13	128057900	C	T	0.00735294	0.992647	0.795455	0.204545	0.870819
13	139037123	C	T	0.0588235	0.941176	0.931818	0.0681818	0.867711
4	68729668	T	A	0.948529	0.0514706	0.0909091	0.909091	0.863824
8	19983547	A	G	0.926471	0.0735294	0.0454545	0.954545	0.859993
7	141920275	C	T	1	0	0.25	0.75	0.858758
7	141920169	CA	C	1	0	0.25	0.75	0.858758
7	141906246	G	A	1	0	0.25	0.75	0.858758
7	141892729	C	T	1	0	0.25	0.75	0.858758
20	6322923	C	T	1	0	0.25	0.75	0.858758
20	6308275	C	T	1	0	0.25	0.75	0.858758
20	6293251	T	C	1	0	0.25	0.75	0.858758
15	12920190	T	C	1	0	0.25	0.75	0.858758
14	77820799	C	A	1	0	0.25	0.75	0.858758
14	77807693	A	T	1	0	0.25	0.75	0.858758
14	77776389	C	T	1	0	0.25	0.75	0.858758
14	77767689	G	A	1	0	0.25	0.75	0.858758
14	29963542	A	C	1	0	0.25	0.75	0.858758
14	29963540	T	G	1	0	0.25	0.75	0.858758
14	29962964	CTACAGA	C	1	0	0.25	0.75	0.858758
14	29961327	C	T	1	0	0.25	0.75	0.858758
14	29960337	T	C	1	0	0.25	0.75	0.858758
14	29960296	TTCA	T	1	0	0.25	0.75	0.858758
14	29958061	G	A	1	0	0.25	0.75	0.858758
14	29950259	T	C	1	0	0.25	0.75	0.858758
14	29946276	C	T	1	0	0.25	0.75	0.858758

14	29945777	T	C	1	0	0.25	0.75	0.858758
12	83383222	T	C	1	0	0.25	0.75	0.858758
12	83383220	G	A	1	0	0.25	0.75	0.858758
12	83379350	T	C	1	0	0.25	0.75	0.858758
7	141928066	C	T	0.963235	0.0367647	0.136364	0.863636	0.858397
7	141920857	T	C	1	0	0.25	0.75	0.858353
7	141888540	A	G	1	0	0.25	0.75	0.858353
4	69069406	T	C	1	0	0.25	0.75	0.858353
20	6320917	T	G	1	0	0.25	0.75	0.858353
14	77805416	G	A	1	0	0.25	0.75	0.858353
14	77796438	A	G	1	0	0.25	0.75	0.858353
14	77795502	A	T	1	0	0.25	0.75	0.858353
14	29945922	C	T	1	0	0.25	0.75	0.858353
14	29943326	G	A	1	0	0.25	0.75	0.858353
14	29943316	A	G	1	0	0.25	0.75	0.858353
12	83379951	A	G	1	0	0.25	0.75	0.858353
12	83374861	TTC	T	1	0	0.25	0.75	0.858353
18	63114822	T	A	1	0	0.25	0.75	0.857947
14	77782171	C	G	1	0	0.25	0.75	0.857947
14	7775072	G	A	1	0	0.25	0.75	0.857947
14	29963329	C	A	1	0	0.25	0.75	0.857947
12	83327373	T	C	1	0	0.25	0.75	0.857947
7	105900502	T	C	1	0	0.25	0.75	0.857541
7	105860669	T	C	1	0	0.25	0.75	0.857541
20	6307739	G	T	0.992647	0.00735294	0.227273	0.772727	0.856827
20	6307711	C	T	0.992647	0.00735294	0.227273	0.772727	0.856827
14	77794630	T	C	0.992647	0.00735294	0.227273	0.772727	0.856827
13	139039755	G	A	0.00735294	0.992647	0.772727	0.227273	0.856827
14	29945376	T	C	0.992647	0.00735294	0.227273	0.772727	0.856438
13	139039720	G	A	0.00735294	0.992647	0.772727	0.227273	0.856438
7	134034132	A	G	0.977941	0.0220588	0.181818	0.818182	0.856074
16	34969816	T	G	0.992647	0.00735294	0.227273	0.772727	0.856049
4	68509885	G	C	0.985294	0.0147059	0.204545	0.795455	0.855974
8	19983951	T	C	0.933824	0.0661765	0.0681818	0.931818	0.855697
8	19622998	T	C	0.933824	0.0661765	0.0681818	0.931818	0.855697
7	141928544	T	C	0.985294	0.0147059	0.204545	0.795455	0.8556
13	128057358	GACCTGCCGAA	G	0.0147059	0.985294	0.795455	0.204545	0.8556
14	29942597	G	A	0.933824	0.0661765	0.0681818	0.931818	0.855393
8	19616794	T	C	0.933824	0.0661765	0.0681818	0.931818	0.854785
2	8424408	T	G	0.911765	0.0882353	0.0227273	0.977273	0.854074
8	19981960	T	C	0.941176	0.0588235	0.0909091	0.909091	0.851594
8	19984063	C	T	0.941176	0.0588235	0.0909091	0.909091	0.85128
14	29963573	G	C	0.941176	0.0588235	0.0909091	0.909091	0.85128
2	8424699	T	A	0.897059	0.102941	0	1	0.849527
7	141927012	T	C	0.955882	0.0441176	0.136364	0.863636	0.845366
4	68503461	A	G	1	0	0.272727	0.727273	0.843909
20	6321945	A	T	1	0	0.272727	0.727273	0.843909
20	6291520	C	CAA	1	0	0.272727	0.727273	0.843909
14	77779676	C	T	1	0	0.272727	0.727273	0.843909

14	29962973	TGGTAC	T	1	0	0.272727	0.727273	0.843909
12	83385359	T	C	1	0	0.272727	0.727273	0.843909
12	83385301	A	C	1	0	0.272727	0.727273	0.843909
12	83381971	G	C	1	0	0.272727	0.727273	0.843909
7	141914176	T	C	1	0	0.272727	0.727273	0.843483
7	141914171	T	G	1	0	0.272727	0.727273	0.843483
7	141895751	T	C	1	0	0.272727	0.727273	0.843483
4	69082103	T	A	1	0	0.272727	0.727273	0.843483
20	6323506	C	T	1	0	0.272727	0.727273	0.843483
20	6318033	A	G	1	0	0.272727	0.727273	0.843483
20	6317925	A	G	1	0	0.272727	0.727273	0.843483
20	6316350	A	T	1	0	0.272727	0.727273	0.843483
15	12923090	T	G	1	0	0.272727	0.727273	0.843483
14	77813691	CT	C	1	0	0.272727	0.727273	0.843483
14	77809048	T	C	1	0	0.272727	0.727273	0.843483
14	77802679	A	G	1	0	0.272727	0.727273	0.843483
14	77799922	G	C	1	0	0.272727	0.727273	0.843483
14	77795446	C	T	1	0	0.272727	0.727273	0.843483
14	77790898	T	C	1	0	0.272727	0.727273	0.843483
14	29962441	T	A	1	0	0.272727	0.727273	0.843483
14	2447634	A	G	1	0	0.272727	0.727273	0.843483
12	83379909	A	G	1	0	0.272727	0.727273	0.843483
12	83379905	T	C	1	0	0.272727	0.727273	0.843483
8	19623187	A	C	0.926471	0.0735294	0.0681818	0.931818	0.843472
7	141923158	T	C	1	0	0.272727	0.727273	0.843057
4	69067888	A	G	1	0	0.272727	0.727273	0.843057
14	83370926	A	G	1	0	0.272727	0.727273	0.843057
14	77793393	A	ACTC	1	0	0.272727	0.727273	0.843057
13	128056521	G	A	0	1	0.727273	0.272727	0.843057
12	83376477	G	A	1	0	0.272727	0.727273	0.843057
12	83327383	C	G	1	0	0.272727	0.727273	0.843057
12	83326665	T	C	1	0	0.272727	0.727273	0.843057
7	105950752	T	G	1	0	0.272727	0.727273	0.842631
7	105949990	G	A	1	0	0.272727	0.727273	0.842631
7	105944658	C	T	1	0	0.272727	0.727273	0.842631
7	105939189	T	C	1	0	0.272727	0.727273	0.842631
7	105909877	A	G	1	0	0.272727	0.727273	0.842631
7	105909515	G	C	1	0	0.272727	0.727273	0.842631
7	105908055	A	T	1	0	0.272727	0.727273	0.842631
7	105907831	G	A	1	0	0.272727	0.727273	0.842631
7	105898532	C	G	1	0	0.272727	0.727273	0.842631
7	105894196	C	T	1	0	0.272727	0.727273	0.842631
7	105853987	A	C	1	0	0.272727	0.727273	0.842631
7	105847428	A	G	1	0	0.272727	0.727273	0.842631
7	105840659	C	T	1	0	0.272727	0.727273	0.842631
7	105764175	A	G	1	0	0.272727	0.727273	0.842631
16	35654879	C	A	0.992647	0.00735294	0.25	0.75	0.841296
14	2450312	T	C	0.992647	0.00735294	0.25	0.75	0.841296
13	128591652	G	A	0.992647	0.00735294	0.25	0.75	0.841296

13	128056118	G	T	0.00735294	0.992647	0.75	0.25	0.841296
13	128056016	A	G	0.00735294	0.992647	0.75	0.25	0.841296
16	34954208	C	T	0.992647	0.00735294	0.25	0.75	0.840888
16	34953776	G	A	0.992647	0.00735294	0.25	0.75	0.840888
16	34951186	A	G	0.992647	0.00735294	0.25	0.75	0.840888
16	34949332	A	G	0.992647	0.00735294	0.25	0.75	0.840888
16	34691501	T	C	0.992647	0.00735294	0.25	0.75	0.840888
14	77803990	T	C	0.992647	0.00735294	0.25	0.75	0.840888
16	34690102	A	G	0.992647	0.00735294	0.25	0.75	0.84048
13	128064962	A	T	0.0147059	0.985294	0.772727	0.227273	0.840247
13	128592571	G	A	0.977941	0.0220588	0.204545	0.795455	0.840202
16	34966327	A	G	0.985294	0.0147059	0.227273	0.772727	0.839856
16	34674286	A	G	0.985294	0.0147059	0.227273	0.772727	0.839856
8	19984032	A	G	0.933824	0.0661765	0.0909091	0.909091	0.839156
4	68740566	TA	T	0.941176	0.0588235	0.113636	0.886364	0.835361
8	19620606	C	G	0.941176	0.0588235	0.113636	0.886364	0.835034
14	29963576	G	C	0.941176	0.0588235	0.113636	0.886364	0.835034
2	8447829	T	C	0.919118	0.0808824	0.0681818	0.931818	0.832261
7	141911843	A	G	1	0	0.295455	0.704545	0.828673
4	68502615	C	T	1	0	0.295455	0.704545	0.828673
20	6323131	A	G	1	0	0.295455	0.704545	0.828673
20	6323127	T	C	1	0	0.295455	0.704545	0.828673
20	6316609	T	A	1	0	0.295455	0.704545	0.828673
20	6316326	A	G	1	0	0.295455	0.704545	0.828673
20	6292940	A	G	1	0	0.295455	0.704545	0.828673
20	6291387	G	A	1	0	0.295455	0.704545	0.828673
20	6291386	T	A	1	0	0.295455	0.704545	0.828673
12	83385691	C	G	1	0	0.295455	0.704545	0.828673
12	83372090	T	G	1	0	0.295455	0.704545	0.828673
7	141920593	G	A	1	0	0.295455	0.704545	0.828225
7	141903301	T	C	1	0	0.295455	0.704545	0.828225
7	141895755	T	C	1	0	0.295455	0.704545	0.828225
7	141888547	A	G	1	0	0.295455	0.704545	0.828225
4	69078454	A	G	1	0	0.295455	0.704545	0.828225
20	6321749	C	T	1	0	0.295455	0.704545	0.828225
20	6320913	A	T	1	0	0.295455	0.704545	0.828225
20	6318190	A	G	1	0	0.295455	0.704545	0.828225
20	6314690	G	T	1	0	0.295455	0.704545	0.828225
20	6313026	A	G	1	0	0.295455	0.704545	0.828225
14	77805119	C	T	1	0	0.295455	0.704545	0.828225
14	77805111	T	C	1	0	0.295455	0.704545	0.828225
14	77804180	A	G	1	0	0.295455	0.704545	0.828225
14	77782939	G	A	1	0	0.295455	0.704545	0.828225
14	148707564	T	G	1	0	0.295455	0.704545	0.828225
14	148705081	C	T	1	0	0.295455	0.704545	0.828225
12	83385156	C	T	1	0	0.295455	0.704545	0.828225
12	83383429	A	G	1	0	0.295455	0.704545	0.828225
12	83376409	A	G	1	0	0.295455	0.704545	0.828225
12	83376400	T	C	1	0	0.295455	0.704545	0.828225

7	141906263	C	T	1	0	0.295455	0.704545	0.827776
7	141901770	A	G	1	0	0.295455	0.704545	0.827776
16	35650255	G	GC	1	0	0.295455	0.704545	0.827776
16	34947890	A	G	1	0	0.295455	0.704545	0.827776
14	83383379	G	A	1	0	0.295455	0.704545	0.827776
14	83383366	C	T	1	0	0.295455	0.704545	0.827776
14	83379697	A	G	1	0	0.295455	0.704545	0.827776
14	83373295	T	C	1	0	0.295455	0.704545	0.827776
14	77786106	A	G	1	0	0.295455	0.704545	0.827776
14	77786028	C	T	1	0	0.295455	0.704545	0.827776
12	83385494	G	A	1	0	0.295455	0.704545	0.827776
12	83376587	C	G	1	0	0.295455	0.704545	0.827776
9	107738309	G	A	1	0	0.295455	0.704545	0.827328
7	105946039	A	G	1	0	0.295455	0.704545	0.827328
7	105898100	T	C	1	0	0.295455	0.704545	0.827328
7	105894805	T	C	1	0	0.295455	0.704545	0.827328
7	105892029	A	G	1	0	0.295455	0.704545	0.827328
7	105880172	T	G	1	0	0.295455	0.704545	0.827328
7	105861998	C	A	1	0	0.295455	0.704545	0.827328
7	105855050	A	G	1	0	0.295455	0.704545	0.827328
7	105849223	T	A	1	0	0.295455	0.704545	0.827328
7	105848525	A	G	1	0	0.295455	0.704545	0.827328
7	105592192	T	C	1	0	0.295455	0.704545	0.827328
16	34952071	C	A	1	0	0.295455	0.704545	0.827328
16	34952069	T	A	1	0	0.295455	0.704545	0.827328
7	105894207	T	C	1	0	0.295455	0.704545	0.82688
7	105862265	T	C	1	0	0.295455	0.704545	0.82688
4	69074757	T	C	0.963235	0.0367647	0.181818	0.818182	0.82618
4	68740595	A	G	0.992647	0.00735294	0.272727	0.727273	0.826177
13	139062328	A	G	0.00735294	0.992647	0.727273	0.272727	0.826177
14	77780781	G	A	0.0955882	0.904412	0.954545	0.0454545	0.825952
7	134034021	G	A	0.970588	0.0294118	0.204545	0.795455	0.825371
16	34950892	A	G	0.992647	0.00735294	0.272727	0.727273	0.825321
13	139051887	A	C	0.00735294	0.992647	0.727273	0.272727	0.825321
13	139043284	A	G	0.00735294	0.992647	0.727273	0.272727	0.825321
13	128055597	C	G	0.00735294	0.992647	0.727273	0.272727	0.825321
4	68507226	T	C	0.985294	0.0147059	0.25	0.75	0.824891
4	68501210	A	G	0.985294	0.0147059	0.25	0.75	0.824891
13	128056945	T	C	0.0147059	0.985294	0.75	0.25	0.824891
7	134035770	T	C	0.977941	0.0220588	0.227273	0.772727	0.824663
4	68509810	T	C	0.985294	0.0147059	0.25	0.75	0.824481
13	128592563	C	T	0.977941	0.0220588	0.227273	0.772727	0.82427
16	34691786	C	A	0.985294	0.0147059	0.25	0.75	0.82407
16	35633947	G	A	0.977941	0.0220588	0.227273	0.772727	0.823876
16	35633931	A	G	0.977941	0.0220588	0.227273	0.772727	0.823876
16	34967005	G	C	0.985294	0.0147059	0.25	0.75	0.82366
2	8449770	T	C	0.933824	0.0661765	0.113636	0.886364	0.822822
15	12905441	C	T	0.933824	0.0661765	0.113636	0.886364	0.822822
8	19997199	A	T	0.933824	0.0661765	0.113636	0.886364	0.822492

8	19997196	TG	T	0.933824	0.0661765	0.113636	0.886364	0.822492
2	8512565	T	A	0.941176	0.0588235	0.136364	0.863636	0.818358
8	19982496	T	C	0.948529	0.0514706	0.159091	0.840909	0.815142
8	19982490	A	G	0.948529	0.0514706	0.159091	0.840909	0.815142
15	12907768	G	T	0.948529	0.0514706	0.159091	0.840909	0.815142
13	139037037	A	C	0.117647	0.882353	0.977273	0.0227273	0.813057
20	6323034	G	A	1	0	0.318182	0.681818	0.813034
20	6320717	T	G	1	0	0.318182	0.681818	0.813034
14	29945950	G	A	1	0	0.318182	0.681818	0.813034
12	83372055	G	A	1	0	0.318182	0.681818	0.813034
7	141924292	A	G	1	0	0.318182	0.681818	0.812562
7	141924291	C	T	1	0	0.318182	0.681818	0.812562
7	141923796	G	A	1	0	0.318182	0.681818	0.812562
7	141911363	C	T	1	0	0.318182	0.681818	0.812562
4	69073241	T	C	1	0	0.318182	0.681818	0.812562
4	69060006	C	T	1	0	0.318182	0.681818	0.812562
4	69056590	G	A	1	0	0.318182	0.681818	0.812562
20	6323087	A	ATCAG	1	0	0.318182	0.681818	0.812562
20	6322483	AC	A	1	0	0.318182	0.681818	0.812562
20	6321933	A	G	1	0	0.318182	0.681818	0.812562
20	6318156	A	T	1	0	0.318182	0.681818	0.812562
20	6316400	T	C	1	0	0.318182	0.681818	0.812562
14	77792564	GTA	G	1	0	0.318182	0.681818	0.812562
14	77787165	T	G	1	0	0.318182	0.681818	0.812562
14	77787164	T	A	1	0	0.318182	0.681818	0.812562
12	83385728	T	C	1	0	0.318182	0.681818	0.812562
12	83385396	A	G	1	0	0.318182	0.681818	0.812562
12	83377304	C	T	1	0	0.318182	0.681818	0.812562
7	141913138	A	G	1	0	0.318182	0.681818	0.81209
7	141904224	C	T	1	0	0.318182	0.681818	0.81209
7	141901779	CTT	C	1	0	0.318182	0.681818	0.81209
7	141899422	A	G	1	0	0.318182	0.681818	0.81209
7	141883762	G	C	1	0	0.318182	0.681818	0.81209
20	6322725	C	T	1	0	0.318182	0.681818	0.81209
16	35653700	A	G	1	0	0.318182	0.681818	0.81209
14	83382217	G	A	1	0	0.318182	0.681818	0.81209
14	83380538	A	G	1	0	0.318182	0.681818	0.81209
14	77804624	C	T	1	0	0.318182	0.681818	0.81209
14	77799862	T	G	1	0	0.318182	0.681818	0.81209
14	77777021	A	G	1	0	0.318182	0.681818	0.81209
14	2449788	G	A	1	0	0.318182	0.681818	0.81209
14	2449777	T	C	1	0	0.318182	0.681818	0.81209
12	83379967	G	A	1	0	0.318182	0.681818	0.81209
12	83369872	G	A	1	0	0.318182	0.681818	0.81209
9	107748721	C	A	1	0	0.318182	0.681818	0.811618
7	105941003	G	A	1	0	0.318182	0.681818	0.811618
7	105896522	A	G	1	0	0.318182	0.681818	0.811618
7	105896398	T	G	1	0	0.318182	0.681818	0.811618
7	105895773	C	G	1	0	0.318182	0.681818	0.811618

7	105890070	A	C	1	0	0.318182	0.681818	0.811618
7	105888252	C	CT	1	0	0.318182	0.681818	0.811618
7	105884212	A	G	1	0	0.318182	0.681818	0.811618
7	105853722	T	A	1	0	0.318182	0.681818	0.811618
7	105852991	T	C	1	0	0.318182	0.681818	0.811618
7	105852720	C	T	1	0	0.318182	0.681818	0.811618
7	105852529	G	A	1	0	0.318182	0.681818	0.811618
7	105850983	T	A	1	0	0.318182	0.681818	0.811618
7	105840514	G	C	1	0	0.318182	0.681818	0.811618
7	105770227	A	G	1	0	0.318182	0.681818	0.811618
7	105760908	T	C	1	0	0.318182	0.681818	0.811618
7	105744584	A	C	1	0	0.318182	0.681818	0.811618
7	105591433	A	G	1	0	0.318182	0.681818	0.811618
14	77804141	G	C	1	0	0.318182	0.681818	0.811618
14	77799755	T	C	1	0	0.318182	0.681818	0.811618
14	77795895	A	G	1	0	0.318182	0.681818	0.811618
14	77793009	A	C	1	0	0.318182	0.681818	0.811618
14	77782087	C	T	1	0	0.318182	0.681818	0.811618
12	83326625	AATTTTTTGAGGTTG AT	A	1	0	0.318182	0.681818	0.811146
2	8410258	G	A	0.926471	0.0735294	0.113636	0.886364	0.8104
12	83316459	T	C	0.992647	0.00735294	0.295455	0.704545	0.810233
8	19999471	G	C	0.963235	0.0367647	0.204545	0.795455	0.809969
4	69070943	A	C	0.992647	0.00735294	0.295455	0.704545	0.809783
16	35654275	T	TTTTA	0.992647	0.00735294	0.295455	0.704545	0.809783
14	29940414	G	A	0.992647	0.00735294	0.295455	0.704545	0.809783
13	139054759	T	TCTC	0.00735294	0.992647	0.704545	0.295455	0.809783
4	69054992	A	G	0.963235	0.0367647	0.204545	0.795455	0.809589
16	34965073	T	A	0.992647	0.00735294	0.295455	0.704545	0.809333
16	34965055	A	G	0.992647	0.00735294	0.295455	0.704545	0.809333
16	34953696	C	T	0.992647	0.00735294	0.295455	0.704545	0.809333
16	34952599	A	C	0.992647	0.00735294	0.295455	0.704545	0.809333
16	34951310	T	C	0.992647	0.00735294	0.295455	0.704545	0.809333
16	34950862	T	G	0.992647	0.00735294	0.295455	0.704545	0.809333
13	139109235	C	T	0.00735294	0.992647	0.704545	0.295455	0.809333
13	139040042	T	C	0.00735294	0.992647	0.704545	0.295455	0.809333
7	105776937	A	G	0.992647	0.00735294	0.295455	0.704545	0.808882
16	34948267	A	G	0.992647	0.00735294	0.295455	0.704545	0.808882
16	34672058	A	ACTTTTT	0.992647	0.00735294	0.295455	0.704545	0.808882
7	141863269	T	C	0.970588	0.0294118	0.227273	0.772727	0.808502
7	134022942	A	T	0.970588	0.0294118	0.227273	0.772727	0.808502
18	26699202	T	G	0.970588	0.0294118	0.227273	0.772727	0.808502
18	26695526	G	C	0.970588	0.0294118	0.227273	0.772727	0.808502
18	26681450	T	C	0.970588	0.0294118	0.227273	0.772727	0.808502
4	68508737	T	G	0.985294	0.0147059	0.272727	0.727273	0.808285
4	68506736	C	A	0.985294	0.0147059	0.272727	0.727273	0.808285
13	139038008	G	C	0.0147059	0.985294	0.727273	0.272727	0.808285
13	128056950	A	C	0.0147059	0.985294	0.727273	0.272727	0.808285
16	35632376	C	T	0.970588	0.0294118	0.227273	0.772727	0.808106

16	34950799	T	C	0.985294	0.0147059	0.272727	0.727273	0.807854
16	34693164	T	C	0.985294	0.0147059	0.272727	0.727273	0.807854
16	34965750	A	G	0.977941	0.0220588	0.25	0.75	0.807491
4	68728987	C	T	0.933824	0.0661765	0.136364	0.863636	0.805388
15	12920378	T	C	0.941176	0.0588235	0.159091	0.840909	0.801237
8	19998235	C	T	0.919118	0.0808824	0.113636	0.886364	0.798095
8	19999037	T	C	0.948529	0.0514706	0.181818	0.818182	0.798038
8	19998021	C	T	0.919118	0.0808824	0.113636	0.886364	0.797427
2	8513351	AC	A	0.948529	0.0514706	0.181818	0.818182	0.7973
20	6315601	C	T	1	0	0.340909	0.659091	0.796978
7	134039896	G	T	1	0	0.340909	0.659091	0.796481
20	6323944	G	A	1	0	0.340909	0.659091	0.796481
20	6323115	C	T	1	0	0.340909	0.659091	0.796481
20	6323039	T	C	1	0	0.340909	0.659091	0.796481
20	6318296	T	G	1	0	0.340909	0.659091	0.796481
20	6290479	G	A	1	0	0.340909	0.659091	0.796481
14	148707609	A	G	1	0	0.340909	0.659091	0.796481
14	148706033	C	T	1	0	0.340909	0.659091	0.796481
12	83377596	C	T	1	0	0.340909	0.659091	0.796481
12	83328395	G	A	1	0	0.340909	0.659091	0.796481
7	141887548	T	C	1	0	0.340909	0.659091	0.795983
20	6312694	A	G	1	0	0.340909	0.659091	0.795983
16	35651190	T	A	1	0	0.340909	0.659091	0.795983
14	77804554	C	T	1	0	0.340909	0.659091	0.795983
14	77783686	G	T	1	0	0.340909	0.659091	0.795983
14	77782990	T	TAG	1	0	0.340909	0.659091	0.795983
14	77781966	C	T	1	0	0.340909	0.659091	0.795983
14	29940458	G	A	1	0	0.340909	0.659091	0.795983
14	2449781	A	G	1	0	0.340909	0.659091	0.795983
14	2235002	G	C	1	0	0.340909	0.659091	0.795983
12	83384104	A	AT	1	0	0.340909	0.659091	0.795983
12	83376405	C	T	1	0	0.340909	0.659091	0.795983
9	107753289	GA	G	1	0	0.340909	0.659091	0.795485
9	107750652	G	A	1	0	0.340909	0.659091	0.795485
9	107749817	A	G	1	0	0.340909	0.659091	0.795485
9	107749816	C	A	1	0	0.340909	0.659091	0.795485
9	107748101	C	T	1	0	0.340909	0.659091	0.795485
9	107745377	T	C	1	0	0.340909	0.659091	0.795485
9	107739286	A	T	1	0	0.340909	0.659091	0.795485
7	141914229	A	G	1	0	0.340909	0.659091	0.795485
7	141914155	GA	G	1	0	0.340909	0.659091	0.795485
7	141901457	C	T	1	0	0.340909	0.659091	0.795485
7	105941888	T	C	1	0	0.340909	0.659091	0.795485
7	105941885	A	G	1	0	0.340909	0.659091	0.795485
7	105890183	A	C	1	0	0.340909	0.659091	0.795485
7	105889855	AC	A	1	0	0.340909	0.659091	0.795485
7	105855428	A	G	1	0	0.340909	0.659091	0.795485
7	105855163	G	C	1	0	0.340909	0.659091	0.795485
7	105854948	A	G	1	0	0.340909	0.659091	0.795485

7	105854845	T	C	1	0	0.340909	0.659091	0.795485
7	105854331	G	A	1	0	0.340909	0.659091	0.795485
7	105854048	A	G	1	0	0.340909	0.659091	0.795485
7	105853524	C	T	1	0	0.340909	0.659091	0.795485
7	105852094	A	G	1	0	0.340909	0.659091	0.795485
7	105849366	TA	T	1	0	0.340909	0.659091	0.795485
7	105847641	T	C	1	0	0.340909	0.659091	0.795485
7	105773386	G	A	1	0	0.340909	0.659091	0.795485
7	105743142	A	G	1	0	0.340909	0.659091	0.795485
7	105740810	T	G	1	0	0.340909	0.659091	0.795485
7	105713469	C	CAG	1	0	0.340909	0.659091	0.795485
7	105712706	G	A	1	0	0.340909	0.659091	0.795485
7	105712699	C	T	1	0	0.340909	0.659091	0.795485
7	105706810	C	T	1	0	0.340909	0.659091	0.795485
7	105702670	G	C	1	0	0.340909	0.659091	0.795485
7	105701193	T	C	1	0	0.340909	0.659091	0.795485
7	105587400	C	A	1	0	0.340909	0.659091	0.795485
7	105585356	T	A	1	0	0.340909	0.659091	0.795485
14	77791742	A	G	1	0	0.340909	0.659091	0.795485
14	77791091	G	T	1	0	0.340909	0.659091	0.795485
14	77788270	A	G	1	0	0.340909	0.659091	0.795485
14	77787671	A	T	1	0	0.340909	0.659091	0.795485
14	77787669	C	T	1	0	0.340909	0.659091	0.795485
14	77786100	G	T	1	0	0.340909	0.659091	0.795485
14	77783643	A	G	1	0	0.340909	0.659091	0.795485
14	77782070	C	T	1	0	0.340909	0.659091	0.795485
14	77774216	T	C	1	0	0.340909	0.659091	0.795485
7	141924066	T	C	1	0	0.340909	0.659091	0.794987
7	105745649	C	T	1	0	0.340909	0.659091	0.794987
12	83386546	C	T	1	0	0.340909	0.659091	0.794987
16	35654520	G	A	0.992647	0.00735294	0.318182	0.681818	0.792908
16	34973685	T	C	0.992647	0.00735294	0.318182	0.681818	0.792908
16	34953686	G	T	0.992647	0.00735294	0.318182	0.681818	0.792908
16	34951697	GA	G	0.992647	0.00735294	0.318182	0.681818	0.792908
16	34951695	GA	G	0.992647	0.00735294	0.318182	0.681818	0.792908
16	34950851	G	A	0.992647	0.00735294	0.318182	0.681818	0.792908
16	34950844	T	TATTTG	0.992647	0.00735294	0.318182	0.681818	0.792908
16	34948666	T	C	0.992647	0.00735294	0.318182	0.681818	0.792908
14	29959699	A	G	0.992647	0.00735294	0.318182	0.681818	0.792908
12	83311816	C	T	0.992647	0.00735294	0.318182	0.681818	0.792908
8	19983776	C	T	0.926471	0.0735294	0.136364	0.863636	0.792892
7	105754220	C	CAT	0.992647	0.00735294	0.318182	0.681818	0.792433
7	105705902	G	A	0.992647	0.00735294	0.318182	0.681818	0.792433
16	34969591	T	TCA	0.992647	0.00735294	0.318182	0.681818	0.792433
16	34952439	C	T	0.992647	0.00735294	0.318182	0.681818	0.792433
16	34949445	C	T	0.992647	0.00735294	0.318182	0.681818	0.792433
13	128056053	A	C	0.00735294	0.992647	0.681818	0.318182	0.792433
12	83360216	T	C	0.867647	0.132353	0.0227273	0.977273	0.791685
4	68507831	C	A	0.985294	0.0147059	0.295455	0.704545	0.791644

18	26696536	T	C	0.970588	0.0294118	0.25	0.75	0.791559
4	68507714	C	T	0.985294	0.0147059	0.295455	0.704545	0.791191
13	128057325	A	G	0.0147059	0.985294	0.704545	0.295455	0.791191
7	134035752	T	TC	0.977941	0.0220588	0.272727	0.727273	0.791086
7	141896509	C	G	0.852941	0.147059	0	1	0.790338
13	128593856	C	T	0.977941	0.0220588	0.272727	0.727273	0.790219
8	19998777	G	A	0.933824	0.0661765	0.159091	0.840909	0.788187
13	139037059	C	T	0.125	0.875	0.954545	0.0454545	0.784517
7	141926379	A	T	0.845588	0.154412	0	1	0.780675
20	6313853	A	G	1	0	0.363636	0.636364	0.780492
20	6308490	C	T	1	0	0.363636	0.636364	0.779966
20	6292091	A	G	1	0	0.363636	0.636364	0.779966
14	148706388	A	G	1	0	0.363636	0.636364	0.779966
20	6320656	A	G	1	0	0.363636	0.636364	0.77944
20	6316480	CT	C	1	0	0.363636	0.636364	0.77944
20	6314113	C	T	1	0	0.363636	0.636364	0.77944
20	6292116	T	C	1	0	0.363636	0.636364	0.77944
14	83378585	A	G	1	0	0.363636	0.636364	0.77944
12	83324243	T	G	1	0	0.363636	0.636364	0.77944
12	83323824	G	A	1	0	0.363636	0.636364	0.77944
12	83316009	A	G	1	0	0.363636	0.636364	0.77944
9	107754825	T	C	1	0	0.363636	0.636364	0.778915
9	107736956	C	T	1	0	0.363636	0.636364	0.778915
9	107735772	G	A	1	0	0.363636	0.636364	0.778915
7	141925143	TTTA	T	1	0	0.363636	0.636364	0.778915
7	141915100	C	T	1	0	0.363636	0.636364	0.778915
7	141914595	A	T	1	0	0.363636	0.636364	0.778915
7	141900857	CCT	C	1	0	0.363636	0.636364	0.778915
7	105949006	CT	C	1	0	0.363636	0.636364	0.778915
7	105911183	C	A	1	0	0.363636	0.636364	0.778915
7	105890461	T	C	1	0	0.363636	0.636364	0.778915
7	105890019	A	G	1	0	0.363636	0.636364	0.778915
7	105889853	G	A	1	0	0.363636	0.636364	0.778915
7	105889847	T	C	1	0	0.363636	0.636364	0.778915
7	105889822	A	G	1	0	0.363636	0.636364	0.778915
7	105887998	T	C	1	0	0.363636	0.636364	0.778915
7	105875343	CCA	C	1	0	0.363636	0.636364	0.778915
7	105854130	T	C	1	0	0.363636	0.636364	0.778915
7	105854060	A	G	1	0	0.363636	0.636364	0.778915
7	105853366	C	T	1	0	0.363636	0.636364	0.778915
7	105853185	A	G	1	0	0.363636	0.636364	0.778915
7	105852398	C	A	1	0	0.363636	0.636364	0.778915
7	105852390	C	A	1	0	0.363636	0.636364	0.778915
7	105848203	C	A	1	0	0.363636	0.636364	0.778915
7	105769849	G	T	1	0	0.363636	0.636364	0.778915
7	105741203	A	T	1	0	0.363636	0.636364	0.778915
7	105741155	A	C	1	0	0.363636	0.636364	0.778915
7	105706400	G	C	1	0	0.363636	0.636364	0.778915
7	105603566	T	C	1	0	0.363636	0.636364	0.778915

7	105592090	G	A	1	0	0.363636	0.636364	0.778915
20	6318276	C	T	1	0	0.363636	0.636364	0.778915
14	77808394	A	C	1	0	0.363636	0.636364	0.778915
14	77805088	C	G	1	0	0.363636	0.636364	0.778915
14	77804060	G	T	1	0	0.363636	0.636364	0.778915
14	77804053	C	T	1	0	0.363636	0.636364	0.778915
14	77784047	C	A	1	0	0.363636	0.636364	0.778915
14	77783824	A	G	1	0	0.363636	0.636364	0.778915
14	77776285	T	A	1	0	0.363636	0.636364	0.778915
14	2235089	T	TGCTGGA	1	0	0.363636	0.636364	0.778915
14	2235032	A	G	1	0	0.363636	0.636364	0.778915
12	83312640	T	A	1	0	0.363636	0.636364	0.778915
7	141901975	C	T	1	0	0.363636	0.636364	0.778388
14	77808487	C	T	1	0	0.363636	0.636364	0.778388
14	77785773	G	A	1	0	0.363636	0.636364	0.778388
14	77779220	G	C	1	0	0.363636	0.636364	0.778388
7	141926449	G	T	0.955882	0.0441176	0.227273	0.772727	0.777984
14	77774357	G	A	1	0	0.363636	0.636364	0.777862
14	2445337	C	T	1	0	0.363636	0.636364	0.777336
4	68740660	C	CTT	0.992647	0.00735294	0.340909	0.659091	0.777028
7	105780466	C	G	1	0	0.363636	0.636364	0.776809
4	69055064	T	G	0.955882	0.0441176	0.227273	0.772727	0.776782
10	6185827	T	C	0.955882	0.0441176	0.227273	0.772727	0.776782
12	83310473	C	CTT	0.992647	0.00735294	0.340909	0.659091	0.776529
16	34968661	A	G	0.992647	0.00735294	0.340909	0.659091	0.776029
16	34967376	C	T	0.992647	0.00735294	0.340909	0.659091	0.776029
16	34953803	G	A	0.992647	0.00735294	0.340909	0.659091	0.776029
16	34950805	G	A	0.992647	0.00735294	0.340909	0.659091	0.776029
16	34948672	A	G	0.992647	0.00735294	0.340909	0.659091	0.776029
14	2448615	A	G	0.992647	0.00735294	0.340909	0.659091	0.776029
12	83329346	A	G	0.992647	0.00735294	0.340909	0.659091	0.776029
12	83311154	T	C	0.992647	0.00735294	0.340909	0.659091	0.776029
7	105709988	C	G	0.992647	0.00735294	0.340909	0.659091	0.775529
7	105603827	A	G	0.992647	0.00735294	0.340909	0.659091	0.775529
16	34951607	A	G	0.992647	0.00735294	0.340909	0.659091	0.775529
16	34950400	A	G	0.992647	0.00735294	0.340909	0.659091	0.775529
16	34949976	C	T	0.992647	0.00735294	0.340909	0.659091	0.775529
13	139053614	T	C	0.00735294	0.992647	0.659091	0.340909	0.775529
18	26681931	A	C	0.963235	0.0367647	0.25	0.75	0.775445
15	12921453	A	G	0.926471	0.0735294	0.159091	0.840909	0.774564
7	134021841	T	C	0.985294	0.0147059	0.318182	0.681818	0.77454
4	68507812	T	C	0.985294	0.0147059	0.318182	0.681818	0.77454
4	68505067	C	A	0.985294	0.0147059	0.318182	0.681818	0.774064
13	139050106	A	G	0.0147059	0.985294	0.681818	0.318182	0.774064
13	128062717	T	C	0.0147059	0.985294	0.681818	0.318182	0.774064
18	26693545	T	C	0.970588	0.0294118	0.272727	0.727273	0.773711
13	128593938	G	A	0.985294	0.0147059	0.318182	0.681818	0.773587
16	34973175	C	T	0.977941	0.0220588	0.295455	0.704545	0.773345
13	128057317	T	C	0.0220588	0.977941	0.704545	0.295455	0.773345

7	105411885	T	C	0.970588	0.0294118	0.272727	0.727273	0.77284
7	105411882	T	C	0.970588	0.0294118	0.272727	0.727273	0.77284
12	83345534	T	A	0.852941	0.147059	0.0227273	0.977273	0.771511
8	19981793	A	C	0.933824	0.0661765	0.181818	0.818182	0.770177
8	19999346	G	A	0.933824	0.0661765	0.181818	0.818182	0.769803
8	19998996	C	A	0.941176	0.0588235	0.204545	0.795455	0.765991
15	12908823	A	G	0.941176	0.0588235	0.204545	0.795455	0.765991
2	8527547	A	C	0.941176	0.0588235	0.204545	0.795455	0.765603
2	8529330	C	T	0.941176	0.0588235	0.204545	0.795455	0.765215
14	148707212	G	C	1	0	0.386364	0.613636	0.763002
14	148706066	G	T	1	0	0.386364	0.613636	0.763002
14	148705925	C	T	1	0	0.386364	0.613636	0.763002
20	6314637	T	C	1	0	0.386364	0.613636	0.762446
20	6292108	T	C	1	0	0.386364	0.613636	0.762446
9	107751807	C	T	1	0	0.386364	0.613636	0.76189
9	107743958	A	G	1	0	0.386364	0.613636	0.76189
9	107735279	C	A	1	0	0.386364	0.613636	0.76189
7	105949022	A	T	1	0	0.386364	0.613636	0.76189
7	105949020	T	G	1	0	0.386364	0.613636	0.76189
7	105907070	A	G	1	0	0.386364	0.613636	0.76189
7	105902193	T	G	1	0	0.386364	0.613636	0.76189
7	105858015	C	T	1	0	0.386364	0.613636	0.76189
7	105854137	T	TC	1	0	0.386364	0.613636	0.76189
7	105849379	C	A	1	0	0.386364	0.613636	0.76189
7	105773954	C	A	1	0	0.386364	0.613636	0.76189
7	105773189	T	C	1	0	0.386364	0.613636	0.76189
7	105770676	G	T	1	0	0.386364	0.613636	0.76189
7	105770630	A	G	1	0	0.386364	0.613636	0.76189
7	105770525	T	C	1	0	0.386364	0.613636	0.76189
7	105766838	A	G	1	0	0.386364	0.613636	0.76189
7	105751202	G	A	1	0	0.386364	0.613636	0.76189
7	105746801	TA	T	1	0	0.386364	0.613636	0.76189
7	105746709	A	G	1	0	0.386364	0.613636	0.76189
7	105746674	C	G	1	0	0.386364	0.613636	0.76189
7	105746321	A	G	1	0	0.386364	0.613636	0.76189
7	105744045	A	C	1	0	0.386364	0.613636	0.76189
7	105592793	G	A	1	0	0.386364	0.613636	0.76189
7	105587126	A	G	1	0	0.386364	0.613636	0.76189
7	105586399	C	CTG	1	0	0.386364	0.613636	0.76189
4	68503036	A	G	1	0	0.386364	0.613636	0.76189
4	68502033	C	T	1	0	0.386364	0.613636	0.76189
20	6316943	A	C	1	0	0.386364	0.613636	0.76189
15	12920079	A	G	1	0	0.386364	0.613636	0.76189
14	83379902	T	C	1	0	0.386364	0.613636	0.76189
14	29942683	C	T	1	0	0.386364	0.613636	0.76189
12	83324071	A	G	1	0	0.386364	0.613636	0.76189
9	107754008	C	T	1	0	0.386364	0.613636	0.761333
7	141914864	G	A	1	0	0.386364	0.613636	0.761333
7	141914806	G	A	1	0	0.386364	0.613636	0.761333

7	141910866	C	T	1	0	0.386364	0.613636	0.761333
7	141896317	A	G	1	0	0.386364	0.613636	0.761333
7	141895646	A	G	1	0	0.386364	0.613636	0.761333
7	141893363	C	T	1	0	0.386364	0.613636	0.761333
7	105888156	A	G	1	0	0.386364	0.613636	0.761333
7	105856801	C	T	1	0	0.386364	0.613636	0.761333
7	105741448	G	A	1	0	0.386364	0.613636	0.761333
7	105697257	T	TAC	1	0	0.386364	0.613636	0.761333
14	77786250	T	C	1	0	0.386364	0.613636	0.761333
14	77780430	T	C	1	0	0.386364	0.613636	0.761333
14	29963570	C	T	1	0	0.386364	0.613636	0.761333
14	29949409	C	A	1	0	0.386364	0.613636	0.761333
12	83376002	TG	T	1	0	0.386364	0.613636	0.761333
12	83322794	G	A	1	0	0.386364	0.613636	0.761333
12	83313891	G	A	1	0	0.386364	0.613636	0.761333
12	83312655	A	T	1	0	0.386364	0.613636	0.761333
2	8424496	T	A	0.919118	0.0808824	0.159091	0.840909	0.761077
7	141924368	C	T	1	0	0.386364	0.613636	0.760777
7	141906016	T	C	1	0	0.386364	0.613636	0.760777
7	141896787	G	C	1	0	0.386364	0.613636	0.760777
14	77795604	C	T	1	0	0.386364	0.613636	0.760777
14	77786979	G	A	1	0	0.386364	0.613636	0.760777
14	77786623	A	T	1	0	0.386364	0.613636	0.760777
14	77778085	C	G	1	0	0.386364	0.613636	0.760777
14	77777659	G	A	1	0	0.386364	0.613636	0.760777
14	29962541	T	C	1	0	0.386364	0.613636	0.760777
14	29957088	T	C	1	0	0.386364	0.613636	0.760777
14	29943138	C	T	1	0	0.386364	0.613636	0.760777
14	2235760	T	C	0.955882	0.0441176	0.25	0.75	0.759976
14	29942204	A	G	0.955882	0.0441176	0.25	0.75	0.759556
14	29942203	C	T	0.955882	0.0441176	0.25	0.75	0.759556
20	6311546	A	G	0.992647	0.00735294	0.363636	0.636364	0.759207
4	69055055	G	A	0.955882	0.0441176	0.25	0.75	0.758717
16	34969559	A	T	0.992647	0.00735294	0.363636	0.636364	0.75868
16	34671751	C	G	0.992647	0.00735294	0.363636	0.636364	0.75868
7	105740568	C	T	0.992647	0.00735294	0.363636	0.636364	0.758152
20	6311569	A	G	0.992647	0.00735294	0.363636	0.636364	0.758152
16	35650108	A	G	0.992647	0.00735294	0.363636	0.636364	0.758152
7	105745524	A	C	0.992647	0.00735294	0.363636	0.636364	0.757624
16	35651901	A	C	0.963235	0.0367647	0.272727	0.727273	0.757463
4	68509897	G	A	0.985294	0.0147059	0.340909	0.659091	0.757459
13	128592054	C	T	0.992647	0.00735294	0.363636	0.636364	0.757096
4	69052195	G	A	0.963235	0.0367647	0.272727	0.727273	0.757025
7	134021883	T	C	0.985294	0.0147059	0.340909	0.659091	0.756957
4	68508232	G	A	0.985294	0.0147059	0.340909	0.659091	0.756455
12	83313019	C	G	0.985294	0.0147059	0.340909	0.659091	0.756455
4	69084696	G	A	0.970588	0.0294118	0.295455	0.704545	0.756247
18	26693699	T	C	0.970588	0.0294118	0.295455	0.704545	0.756247
14	148708541	C	A	0.970588	0.0294118	0.295455	0.704545	0.756247

18	26697017	A	G	0.977941	0.0220588	0.318182	0.681818	0.75603
13	128060694	A	G	0.0147059	0.985294	0.659091	0.340909	0.755953
12	83385407	T	C	0.926471	0.0735294	0.181818	0.818182	0.75573
13	139037894	A	C	0.0220588	0.977941	0.681818	0.318182	0.755551
2	8449905	A	G	0.904412	0.0955882	0.136364	0.863636	0.755139
7	105412444	C	T	0.970588	0.0294118	0.295455	0.704545	0.754418
7	141895123	G	A	0.823529	0.176471	0	1	0.751796
8	19998834	C	T	0.933824	0.0661765	0.204545	0.795455	0.751686
8	19998775	C	T	0.933824	0.0661765	0.204545	0.795455	0.751686

M4. TRANZKRIPCIÓS FAKTOR KÖTŐHELYEK A SZELEKTÁLT RÉGIÓKBAN

Kromoszóm a	Start	Vég	TFBS típus	Kromoszóm a	Start	Vég	TFBS típus
8	19620207	19620219	Hnf4a	13	128590400	128590412	SP1
4	69056129	69056141	Hnf4a	13	139059489	139059501	SP1
4	69082436	69082448	Hnf4a	13	139059467	139059479	SP1
4	69082459	69082471	Hnf4a	13	139059443	139059455	SP1
20	6323797	6323809	Hnf4a	13	139057429	139057441	SP1
9	107740840	10774085 2	Hnf4a	13	139052484	139052496	SP1
7	105840925	10584093 7	Hnf4a	13	139051443	139051455	SP1
4	69082438	69082450	HNF4G	7	141898108	141898120	SP1
7	105840927	10584093 9	HNF4G	7	141903577	141903589	SP1
4	69059595	69059605	HOXA13	14	83379224	83379236	SP1
18	26692874	26692884	JUN	15	12907997	12908009	SP1
4	69061390	69061400	JUN	16	35650362	35650374	SPI1
7	105843432	10584344 2	JUN	16	34975249	34975261	SPI1
4	69061389	69061401	Atf3	16	34975461	34975473	SPI1
4	69073887	69073899	Atf3	16	34978768	34978780	SPI1
7	105843428	10584344 0	Atf3	16	34971880	34971892	SPI1
18	26696135	26696147	JUND	13	139060218	139060230	SPI1
18	26692875	26692887	JUND	13	139053725	139053737	SPI1
4	69057475	69057487	JUND	13	139052482	139052494	SPI1
4	69061391	69061403	JUND	13	139052065	139052077	SPI1
7	105843429	10584344 1	JUND	13	139049467	139049479	SPI1
20	6294104	6294112	KLF4	7	105876914	105876926	SPI1
7	105845076	10584508 4	KLF4	7	105889007	105889019	SPI1
7	105845545	10584555 3	KLF4	7	141878079	141878091	SPI1
7	105845073	10584508 6	KLF9	7	141898110	141898122	SPI1
7	105845645	10584565 7	Mafb	7	141903379	141903391	SPI1
2	9523005	9523017	Mafb	14	77791146	77791158	SPI1
9	107746480	10774649 2	ATF4	7	141928386	141928394	Sreb1(var.2)
7	134038357	13403836 9	ATF4	16	34977758	34977768	SRF
4	69080335	69080354	MAFK	2	8414535	8414545	SRF
9	107745608	10774562 7	MAFK	13	128592776	128592786	STAT1
7	134022337	13402235 6	MAFK	13	128590362	128590372	STAT1
4	69082370	69082382	MAX	16	34978767	34978779	STAT1::STAT2
9	107740799	10774081 1	MAX	13	128592784	128592796	STAT1::STAT2
7	105844649	10584466 1	MAX	13	139073920	139073932	STAT1::STAT2
7	134022319	13402233 1	MAX	4	68508489	68508501	STAT1::STAT2

2	9525455	9525467	MAX	7	141903378	141903390	STAT1::STAT2
7	134022319	134022329	MAX::MYC	16	34673407	34673417	STAT3
4	69073900	69073908	MYB	13	128590363	128590373	STAT3
4	69061396	69061406	MYC	13	139104242	139104252	STAT3
7	105844634	105844644	MYC	7	105899862	105899872	STAT3
7	105844910	105844920	MYC	7	141865488	141865498	STAT3
7	105853750	105853760	MYC	7	141866479	141866489	STAT3
7	134022320	134022330	MYC	7	141897915	141897925	STAT3
4	69061395	69061407	MYCN	14	77791057	77791067	STAT3
7	105844909	105844921	MYCN	13	139106084	139106096	CDX2
7	105853749	105853761	MYCN	13	128592775	128592787	Stat5a::Stat5b
7	134022319	134022331	MYCN	4	68507907	68507919	Stat5a::Stat5b
20	6318326	6318336	NEUROD1	14	77791260	77791270	TAL1::TCF3
7	105843412	105843422	NFIA	13	139053629	139053639	TAL1::TCF3
7	105853727	105853737	NFIA	13	139063895	139063905	TAL1::TCF3
18	26692730	26692740	NFIC	7	141874356	141874366	TAL1::TCF3
8	20036348	20036358	NFIC	7	141866211	141866221	TAL1::TCF3
8	20036130	20036140	NFIC	13	128064787	128064797	TAL1::TCF3
20	6318287	6318297	NFIC	13	139059343	139059353	TAL1::TCF3
9	107746103	107746113	NFIC	13	128595542	128595552	TAL1::TCF3
7	105589382	105589392	NFIC	13	139051995	139052005	TAL1::TCF3
7	105843411	105843421	NFIC	13	128595571	128595581	TAL1::TCF3
7	105845317	105845327	NFIC	13	128594836	128594846	TAL1::TCF3
7	105853748	105853758	NFIC	7	141899963	141899973	TAL1::TCF3
4	69059984	69059994	NFKB2	7	141897204	141897214	TAL1::TCF3
9	107740968	107740980	NFYA	14	29941582	29941592	TAL1::TCF3
9	107740972	107740984	NFYB	13	128065330	128065340	TAL1::TCF3
4	69059567	69059575	Nkx3-1	13	139053657	139053667	TAL1::TCF3
7	105586133	105586141	NKX3-2	7	141866168	141866178	TAL1::TCF3
7	105840926	105840938	NR1H2::RXRA	13	128064892	128064902	TAL1::TCF3
8	20036282	20036294	NR2F1	13	139062884	139062894	TBX1
4	69057381	69057393	NR2F1	7	141928509	141928519	TBX1
4	69064008	69064020	NR2F1	16	34977645	34977657	Tcf12
20	6317946	6317958	NR2F1	13	128065330	128065342	Tcf12
7	105859618	105859630	NR2F1	13	128064892	128064904	Tcf12
2	9528577	9528589	NR2F1	13	139063895	139063907	Tcf12
2	9523976	9523988	NR3C2	13	139059343	139059355	Tcf12
18	26692873	26692885	BATF::JUN	13	139052749	139052761	Tcf12
4	69057474	69057486	BATF::JUN	13	139051995	139052007	Tcf12
4	69061389	69061401	BATF::JUN	13	139050738	139050750	Tcf12
4	69073887	69073899	BATF::JUN	4	68500618	68500630	Tcf12

7	105843428	105843440	BATF::JUN	7	141866166	141866178	Tcf12
2	9527805	9527815	OTX2	7	141874354	141874366	Tcf12
18	26696136	26696146	PBX1	7	141874609	141874621	Tcf12
18	26692835	26692845	PBX1	7	141881555	141881567	Tcf12
2	9525419	9525427	PDX1	14	77791258	77791270	Tcf12
20	6324610	6324622	Bcl6	14	83375843	83375855	Tcf12
18	26696255	26696267	Pou5f1::Sox2	13	139063894	139063906	Tcf21
9	107746558	107746570	Pou5f1::Sox2	7	141866210	141866222	Tcf21
7	105841522	105841534	Pou5f1::Sox2	14	29941581	29941593	Tcf21
4	69059984	69059994	REL	13	128595542	128595552	TCF3
7	105844800	105844812	RUNX1	13	128064892	128064902	TCF3
4	69060623	69060633	RUNX3	13	139063895	139063905	TCF3
7	105585820	105585832	Rxra	13	139059343	139059353	TCF3
7	105840926	105840938	Rxra	13	139053657	139053667	TCF3
4	69057391	69057404	SMAD2::SMAD3::SMAD4	13	139053629	139053639	TCF3
20	6320027	6320040	SMAD2::SMAD3::SMAD4	13	139051995	139052005	TCF3
7	105845338	105845350	BHLHE41	7	141864778	141864788	TCF3
7	105845084	105845096	SNAI2	7	141866168	141866178	TCF3
4	69058687	69058699	Sox2	7	141866211	141866221	TCF3
4	69061484	69061496	Sox2	7	141874356	141874366	TCF3
20	6286069	6286081	Sox2	7	141899963	141899973	TCF3
2	9522722	9522734	Sox2	14	29941582	29941592	TCF3
2	9529808	9529820	Sox2	14	77791260	77791270	TCF3
4	69076826	69076836	SOX9	16	34945127	34945137	CEBPA
4	69082386	69082396	SOX9	13	139055569	139055579	CEBPA
7	105845386	105845398	SP1	4	68725956	68725966	CEBPA
7	105845612	105845624	SP1	7	141874145	141874155	CEBPA
7	105845726	105845738	SP1	12	83311549	83311559	CEBPA
20	6311518	6311530	SPI1	13	128595571	128595581	TCF4
20	6294290	6294302	SPI1	13	128595542	128595552	TCF4
7	105586187	105586199	SPI1	13	128594836	128594846	TCF4
7	105858117	105858129	SPI1	13	128064892	128064902	TCF4
7	134038350	134038362	SPI1	13	128064787	128064797	TCF4
7	105845666	105845676	STAT1	13	139063895	139063905	TCF4
4	69059933	69059943	STAT3	13	139053629	139053639	TCF4
4	69059593	69059605	CDX2	13	139051995	139052005	TCF4
4	69074504	69074516	CDX2	7	105888975	105888985	TCF4
9	107745952	107745964	Stat5a::Stat5b	7	141866168	141866178	TCF4
7	105845665	105845677	Stat5a::Stat5b	7	141874356	141874366	TCF4
20	6324695	6324705	TAL1::TCF3	14	29941582	29941592	TCF4
7	105845339	105845349	TAL1::TCF3	14	77791260	77791270	TCF4

8	20033203	20033213	TAL1::TCF3	14	77769382	77769394	Tcf7
20	6324695	6324707	Tcf12	14	77769383	77769395	TCF7L2
7	105845339	10584535 1	Tcf12	14	83370321	83370333	TCF7L2
2	9520021	9520033	Tcf12	4	68746790	68746800	TEAD1
8	20033202	20033214	Tcf21	14	29949950	29949960	TEAD1
8	20033203	20033213	TCF3	14	29949992	29950002	TEAD1
7	105845339	10584534 9	TCF3	15	12909359	12909369	TEAD1
20	6324695	6324705	TCF4	4	68735442	68735452	TEAD4
20	6323732	6323742	TCF4	4	68746790	68746800	TEAD4
9	107746577	10774658 7	TCF4	7	141896393	141896403	TEAD4
7	105844910	10584492 0	TCF4	14	29949992	29950002	TEAD4
7	105845339	10584534 9	TCF4	12	83358036	83358046	TEAD4
2	9523733	9523743	TEAD1	13	139056925	139056937	TFAP2A
8	20036209	20036219	TEAD4	13	139050704	139050716	TFAP2A
2	9527827	9527837	TEAD4	13	139045734	139045746	TFAP2A
4	69061379	69061391	TFAP2A	13	139035509	139035521	TFAP2A
20	6294267	6294279	TFAP2A	4	68735322	68735334	TFAP2A
18	26692923	26692935	CEBPB	14	77781107	77781119	TFAP2A
4	69060036	69060048	CEBPB	14	83375470	83375482	TFAP2A
4	69070368	69070380	CEBPB	13	139050705	139050717	TFAP2A-var.2
4	69082477	69082489	CEBPB	13	139045735	139045747	TFAP2A-var.2
9	107740983	10774099 5	CEBPB	13	139035508	139035521	TFAP2A-var.3
9	107745908	10774592 0	CEBPB	16	34967305	34967317	CEBPB
9	107746479	10774649 1	CEBPB	16	34945145	34945157	CEBPB
7	134038335	13403834 7	CEBPB	16	34945126	34945138	CEBPB
2	9529259	9529271	CEBPB	16	34942462	34942474	CEBPB
10	6188707	6188719	CEBPB	16	34692864	34692876	CEBPB
8	19984059	19984071	TP63	13	128067982	128067994	CEBPB
7	105841551	10584156 3	TP63	13	128065733	128065745	CEBPB
4	69061396	69061406	USF1	13	139084742	139084754	CEBPB
2	9525467	9525477	USF1	13	139083762	139083774	CEBPB
4	69061396	69061404	USF2	13	139074275	139074287	CEBPB
2	9525458	9525466	USF2	13	139063740	139063752	CEBPB
7	105845252	10584526 0	ZBTB7A	13	139055568	139055580	CEBPB
4	69082453	69082463	ZEB1	13	139053589	139053601	CEBPB
7	105853750	10585376 0	ZEB1	4	68497255	68497267	CEBPB
7	105845620	10584563 2	ZNF143	4	68499283	68499295	CEBPB
7	105845398	10584541 0	ZNF740	4	68725955	68725967	CEBPB
8	20036250	20036260	CREB1	4	68738858	68738870	CEBPB
7	105845105	10584511 7	CTCF	7	105707911	105707923	CEBPB
7	105853750	10585376 2	CTCF	7	105771079	105771091	CEBPB
2	9525538	9525550	CTCF	7	105899276	105899288	CEBPB
20	6319858	6319870	Dux	7	105908197	105908209	CEBPB

20	6283470	6283482	Dux	7	105911367	105911379	CEBPB
4	69074115	69074127	DUX4	7	105913466	105913478	CEBPB
20	6319856	6319868	DUX4	7	141874144	141874156	CEBPB
2	9523715	9523727	DUX4	7	141898128	141898140	CEBPB
8	20036382	20036394	Ar	14	29955027	29955039	CEBPB
4	69074348	69074360	Ar	14	77821061	77821073	CEBPB
20	6319791	6319803	Ar	12	83342560	83342572	CEBPB
20	6292915	6292927	Ar	13	128590390	128590402	TFAP2C
7	105585528	105585540	Ar	13	128065563	128065575	TFAP2C
7	105585591	105585603	Ar	13	139056925	139056937	TFAP2C
20	6309736	6309748	EBF1	13	139051465	139051477	TFAP2C
4	69076640	69076652	ELF1	13	139041056	139041068	TFAP2C
7	134038786	134038798	ELF3	13	139035955	139035967	TFAP2C
2	9525571	9525583	ELF3	13	139035630	139035642	TFAP2C
9	107741081	107741091	ELK4	4	68735322	68735334	TFAP2C
4	69057172	69057184	ERG	7	141877732	141877744	TFAP2C
4	69076641	69076653	ERG	13	139082368	139082380	TFAP4
7	105586185	105586197	ERG	13	139073862	139073874	TFAP4
7	134038348	134038360	ERG	13	139052412	139052424	TFAP4
2	9525572	9525584	ERG	7	105884396	105884408	TP53
4	69057384	69057394	Esrra	7	105884396	105884408	TP63
4	69076650	69076660	Esrra	7	141896500	141896512	TP63
4	69057172	69057182	Esrrg	7	141898419	141898431	TP63
4	69059973	69059983	Esrrg	14	29953837	29953849	TP63
4	69076641	69076651	Esrrg	14	77764570	77764582	TP63
7	105586187	105586197	Esrrg	14	77802149	77802161	TP63
7	134038350	134038360	Esrrg	14	77802239	77802251	TP63
2	9525572	9525582	Esrrg	15	12907937	12907949	TP63
18	26693183	26693195	FLI1	13	139053590	139053600	CEBPD
7	105845636	105845648	FLI1	13	139063489	139063499	USF1
18	26696135	26696147	FOS	4	68502558	68502568	USF1
18	26692875	26692887	FOS	7	105766252	105766262	USF1
4	69057475	69057487	FOS	7	105902237	105902247	USF1
4	69059607	69059619	FOS	7	141897283	141897293	USF1
4	69059804	69059816	FOS	14	2236036	2236046	USF1
4	69061390	69061402	FOS	14	2450154	2450164	USF1
20	6311536	6311548	FOS	15	12909620	12909630	USF1
20	6293324	6293336	FOS	12	83331305	83331315	USF1
20	6286097	6286109	FOS	13	139063491	139063499	USF2
7	105843429	105843441	FOS	14	83381858	83381866	USF2
2	9529909	9529921	FOS	7	141897283	141897291	USF2
20	6293393	6293401	Arnt	13	139063600	139063608	Vdr
18	26692874	26692886	FOS::JUN	13	128064629	128064641	YY1
4	69061390	69061402	FOS::JUN	13	139035400	139035412	YY1
7	105843430	105843442	FOS::JUN	7	105771259	105771271	YY1

18	26696135	26696147	FOSL1	7	105902352	105902364	YY1
4	69057476	69057488	FOSL1	7	141864539	141864551	YY1
4	69057512	69057524	FOSL1	7	141914143	141914155	YY1
4	69061390	69061402	FOSL1	7	141925358	141925370	YY1
7	105843429	10584344 1	FOSL1	7	141928206	141928218	YY1
18	26696135	26696147	FOSL2	7	105899845	105899853	ZBTB7A
18	26692875	26692887	FOSL2	7	141865842	141865850	ZBTB7A
4	69057475	69057487	FOSL2	7	141878079	141878087	ZBTB7A
4	69057513	69057525	FOSL2	14	83370150	83370158	ZBTB7A
4	69061390	69061402	FOSL2	13	128595929	128595939	ZEB1
20	6293324	6293336	FOSL2	13	139054323	139054333	ZEB1
7	105843429	10584344 1	FOSL2	13	139051419	139051429	ZEB1
18	26699450	26699462	FOXA1	4	68737275	68737285	ZEB1
8	20039634	20039646	FOXA1	7	141892655	141892665	ZEB1
8	20036378	20036390	FOXA1	13	128590666	128590678	ZNF143
8	20033341	20033353	FOXA1	13	139061958	139061970	ZNF143
8	19618290	19618302	FOXA1	13	139035212	139035224	ZNF143
8	19616459	19616471	FOXA1	4	68732959	68732971	ZNF143
4	69082701	69082713	FOXA1	14	77813124	77813136	ZNF143
20	6320041	6320053	FOXA1	14	77819261	77819273	ZNF143
9	107752553	10775256 5	FOXA1	13	139058968	139058980	ZNF740
9	107752583	10775259 5	FOXA1	13	139058903	139058915	ZNF740
14	148716474	14871648 6	FOXA1	7	141883256	141883268	ZNF740
7	105585546	10558555 8	FOXA1	7	141928459	141928469	CREB1
7	105585643	10558565 5	FOXA1	14	77778033	77778043	CREB1
7	105841031	10584104 3	FOXA1	14	77791392	77791402	CREB1
7	105841482	10584149 4	FOXA1	14	77819714	77819724	CREB1
7	134034285	13403429 7	FOXA1	16	34968664	34968676	CTCF
2	9529826	9529838	FOXA1	16	34693650	34693662	CTCF
18	26692942	26692954	Foxa2	16	34673369	34673381	CTCF
8	20036375	20036387	Foxa2	13	128594897	128594909	CTCF
8	19616456	19616468	Foxa2	13	139054968	139054980	CTCF
4	69057594	69057606	Foxa2	4	68497209	68497221	CTCF
20	6320044	6320056	Foxa2	2	97360703	97360715	CTCF
7	105841479	10584149 1	Foxa2	7	141864776	141864788	CTCF
7	134034282	13403429 4	Foxa2	7	141866168	141866180	CTCF
7	134035462	13403547 4	Foxa2	7	141903651	141903663	CTCF
2	9529829	9529841	Foxa2	7	141925361	141925373	CTCF
20	6293393	6293401	ARNT::HIF1A	14	77815388	77815400	CTCF
8	19616461	19616471	Foxo1	14	83381867	83381879	CTCF
4	69053073	69053083	Foxo1	12	83311593	83311605	CTCF
4	69076643	69076653	Gabpa	7	105713967	105713979	Dux
2	9525574	9525584	Gabpa	7	141895594	141895606	Dux
4	69061178	69061188	Gata1	14	77800234	77800246	Dux

4	69082633	69082643	Gata1	7	105713965	105713977	DUX4
20	6317767	6317777	Gata1	16	35634919	35634931	Ar
7	105843715	105843725	Gata1	16	34972081	34972093	Ar
8	19620076	19620086	GATA2	13	128063099	128063111	Ar
4	69082633	69082643	GATA2	13	139098711	139098723	Ar
4	69058428	69058440	GATA3	13	139078311	139078323	Ar
4	69074277	69074289	GATA3	13	139064509	139064521	Ar
20	6292928	6292940	GATA3	7	105413819	105413831	Ar
7	105590391	105590403	GATA3	7	141865156	141865168	Ar
7	105591309	105591321	GATA3	7	141866187	141866199	Ar
2	9523278	9523290	GATA3	7	141866978	141866990	Ar
8	20039769	20039781	Gata4	7	141874529	141874541	Ar
8	19620077	19620089	Gata4	14	29952092	29952104	Ar
4	69058430	69058442	Gata4	14	2447992	2448004	Ar
20	6317768	6317780	Gata4	14	77791294	77791306	Ar
8	19620075	19620087	GATA5	14	83380778	83380790	Ar
4	69058428	69058440	GATA5	15	12909880	12909892	Ar
4	69076617	69076629	GRHL1	13	139057229	139057239	E2F6
13	128596256	128596266	HEY1	7	141928432	141928442	E2F6
13	139061712	139061724	HNF1B	13	128595465	128595477	EBF1
13	128593955	128593967	Hnf4a	13	139040572	139040584	EBF1
13	139038888	139038900	Hnf4a	7	141886067	141886079	EBF1
7	141865303	141865315	Hnf4a	14	29941695	29941707	EBF1
7	141896387	141896399	Hnf4a	14	2445326	2445338	EBF1
7	141896489	141896501	Hnf4a	13	139059150	139059162	EGR1
14	77802951	77802963	Hnf4a	13	139057426	139057438	EGR1
14	83378622	83378634	Hnf4a	13	139051966	139051978	EGR1
7	141892633	141892645	HNF4G	7	141864285	141864297	EGR1
7	141864813	141864825	HSF1	14	77791126	77791138	EGR1
7	141874192	141874204	HSF1	14	77817437	77817449	EGR1
4	68497268	68497280	IRF1	13	139063969	139063981	ELF1
13	128064758	128064768	JUN	13	139052483	139052495	ELF1
13	139106535	139106545	JUN	13	139051444	139051456	ELF1
13	139062843	139062853	JUN	4	68735269	68735281	ELF1
13	139058849	139058859	JUN	7	105911327	105911339	ELF1
13	139055226	139055236	JUN	7	141865873	141865885	ELF1
13	139038835	139038845	JUN	7	141884974	141884986	ELF1
4	68508860	68508870	JUN	7	141926048	141926060	ELF1
4	68749616	68749626	JUN	14	29942029	29942041	ELF1
7	141898101	141898111	JUN	4	68497150	68497158	Arid3a
14	2447899	2447909	JUN	7	141865000	141865008	Arid3a

13	139055224	13905523 6	JUNB	14	83370123	83370131	Arid3a
7	141898101	14189811 3	JUNB	13	128592784	128592796	ELF3
13	128590438	12859045 0	Atf3	13	128065623	128065635	ELF3
13	139106537	13910654 9	Atf3	13	139044623	139044635	ELF3
13	139062842	13906285 4	Atf3	4	68508559	68508571	ELF3
13	139055225	13905523 7	Atf3	4	68735269	68735281	ELF3
7	141884682	14188469 4	Atf3	7	105911327	105911339	ELF3
7	141886032	14188604 4	Atf3	7	141884974	141884986	ELF3
7	141896466	14189647 8	Atf3	7	141926214	141926226	ELF3
7	141898100	14189811 2	Atf3	14	77780933	77780945	ELF3
13	128590440	12859045 2	JUND	13	139051444	139051454	ELK1
13	139106536	13910654 8	JUND	12	83351712	83351722	ELK1
13	139055224	13905523 6	JUND	12	83351711	83351721	ELK4
4	68508860	68508872	JUND	16	34969697	34969710	EOMES
7	105745165	10574517 7	JUND	13	128065251	128065264	EOMES
7	105750137	10575014 9	JUND	7	141865347	141865357	ERF
7	141898101	14189811 3	JUND	7	141865486	141865496	ERF
14	2447896	2447908	JUND	7	141898109	141898119	ERF
13	139040763	13904077 3	Klf1	16	34978766	34978778	ERG
16	35650079	35650087	KLF4	13	128592785	128592797	ERG
13	139062823	13906283 1	KLF4	13	128064919	128064931	ERG
13	139057860	13905786 8	KLF4	13	139073919	139073931	ERG
4	68749681	68749689	KLF4	13	139063968	139063980	ERG
13	139059449	13905946 1	KLF5	13	139062921	139062933	ERG
13	139059162	13905917 5	KLF9	13	139059387	139059399	ERG
16	34670819	34670831	Mafb	13	139053295	139053307	ERG
13	128065300	12806531 2	Mafb	13	139053240	139053252	ERG
13	139106540	13910655 2	Mafb	13	139051443	139051455	ERG
4	68732308	68732320	Mafb	7	105711731	105711743	ERG
7	141881639	14188165 1	Mafb	7	141865346	141865358	ERG
16	34692863	34692875	ATF4	7	141884896	141884908	ERG
14	77818960	77818972	ATF4	7	141884975	141884987	ERG
2	8430944	8430956	ATF4	7	141898108	141898120	ERG
2	8448195	8448207	ATF4	7	141928467	141928479	ERG
13	139106536	13910654 8	MAFF	14	29944467	29944479	ERG
7	105701388	10570140 0	MAFF	15	12907960	12907972	ERG
7	141896465	14189647 7	MAFF	12	83373535	83373547	ERG
15	12909413	12909425	MAFF	4	68500589	68500606	ESR1
13	139106536	13910654 8	MAFG	13	128593950	128593960	Esrra

13	139055224	13905523 6	MAFG	13	128061303	128061313	Esrra
7	141870696	14187070 8	MAFG	13	139058668	139058678	Esrra
13	139106538	13910655 0	MAFG::NFE2L1	13	139051469	139051479	Esrra
13	139106532	13910655 1	MAFK	4	68500586	68500596	Esrra
4	68746778	68746797	MAFK	7	141865754	141865764	Esrra
7	105701384	10570140 3	MAFK	7	141898242	141898252	Esrra
7	141870693	14187071 2	MAFK	14	77781149	77781159	Esrra
14	77793020	77793039	MAFK	14	77817422	77817432	Esrra
15	12909409	12909428	MAFK	16	34978768	34978778	Esrrg
13	139106538	13910655 0	MAF::NFE2	13	128592785	128592795	Esrrg
16	34969862	34969874	MAX	13	128064921	128064931	Esrrg
13	139060861	13906087 3	MAX	13	139073921	139073931	Esrrg
13	139055196	13905520 8	MAX	13	139063970	139063980	Esrrg
13	139048225	13904823 7	MAX	13	139059389	139059399	Esrrg
13	139048164	13904817 6	MAX	13	139053331	139053341	Esrrg
14	83381857	83381869	MAX	13	139053240	139053250	Esrrg
7	105888925	10588893 7	MAX	13	139052009	139052019	Esrrg
7	141892007	14189201 9	MAX	13	139051445	139051455	Esrrg
7	141908282	14190829 4	MAX	7	105711733	105711743	Esrrg
7	141925943	14192595 5	MAX	7	141865346	141865356	Esrrg
7	141929610	14192962 2	MAX	7	141878079	141878089	Esrrg
14	29942077	29942089	MAX	7	141884898	141884908	Esrrg
14	2236035	2236047	MAX	7	141884975	141884985	Esrrg
14	2450151	2450163	MAX	7	141898110	141898120	Esrrg
14	77822455	77822467	MAX	7	141928467	141928477	Esrrg
14	83370427	83370439	MAX	14	29944467	29944477	Esrrg
14	83370516	83370528	MAX	15	12907962	12907972	Esrrg
15	12909634	12909646	MAX	13	128592786	128592796	ETS1
12	83339889	83339901	MAX	13	139053241	139053251	ETS1
13	139060861	13906087 1	MAX::MYC	13	139051444	139051454	ETS1
14	83381859	83381869	MAX::MYC	7	141865887	141865897	ETS1
7	141908284	14190829 4	MAX::MYC	7	141884897	141884907	ETS1
14	77805757	77805767	MAX::MYC	7	141898109	141898119	ETS1
14	77822455	77822465	MAX::MYC	7	141898108	141898120	ETV1
14	83370427	83370437	MAX::MYC	13	139051444	139051454	ETV6
15	12909636	12909646	MAX::MYC	7	141865887	141865897	ETV6
12	83339891	83339901	MAX::MYC	7	141878032	141878042	ETV6
4	68508526	68508538	MEF2A	7	141898109	141898119	ETV6
7	141898003	14189801 5	MEF2A	7	141928469	141928477	FEV
7	141929246	14192925 8	MEF2A	13	128064919	128064931	FLI1
13	139060710	13906072 2	MEF2B	13	128056193	128056205	FLI1

4	68508526	68508538	MEF2B	13	139059387	139059399	FLI1
7	141898003	141898015	MEF2B	13	139053240	139053252	FLI1
2	8423609	8423621	MEF2B	7	141884896	141884908	FLI1
13	128591532	128591544	MYBL1	14	29944467	29944479	FLI1
13	139035113	139035125	MYBL1	16	34690366	34690378	FOS
7	141884943	141884955	MYBL1	13	128590440	128590452	FOS
7	141928413	141928425	MYBL1	13	128064756	128064768	FOS
14	29941595	29941607	MYBL1	13	139106536	139106548	FOS
16	34941931	34941941	MYC	13	139055224	139055236	FOS
13	139063489	139063499	MYC	13	139052320	139052332	FOS
13	139060862	139060872	MYC	13	139038833	139038845	FOS
4	68497083	68497093	MYC	4	68508860	68508872	FOS
4	68508512	68508522	MYC	4	68736660	68736672	FOS
14	83381858	83381868	MYC	4	68749616	68749628	FOS
7	141864778	141864788	MYC	7	105745166	105745178	FOS
7	141866168	141866178	MYC	7	105750136	105750148	FOS
7	141866211	141866221	MYC	7	141870696	141870708	FOS
7	141874609	141874619	MYC	7	141884683	141884695	FOS
7	141888615	141888625	MYC	7	141896465	141896477	FOS
7	141895728	141895738	MYC	7	141896795	141896807	FOS
7	141908283	141908293	MYC	7	141897731	141897743	FOS
14	2236036	2236046	MYC	7	141898101	141898113	FOS
14	77805758	77805768	MYC	14	2447896	2447908	FOS
14	77822456	77822466	MYC	14	77768272	77768284	FOS
14	83370873	83370883	MYC	14	77817484	77817496	FOS
14	83376218	83376228	MYC	2	8513559	8513571	FOS
15	12909635	12909645	MYC	13	128590825	128590833	Arnt
16	35651075	35651087	MYCN	13	128065322	128065330	Arnt
13	139063488	139063500	MYCN	13	128064756	128064768	FOS::JUN
13	139060861	139060873	MYCN	13	139106535	139106547	FOS::JUN
4	68508511	68508523	MYCN	13	139055223	139055235	FOS::JUN
14	83381857	83381869	MYCN	13	139038833	139038845	FOS::JUN
7	141908282	141908294	MYCN	4	68508861	68508873	FOS::JUN
14	29941677	29941689	MYCN	4	68749616	68749628	FOS::JUN
14	2236035	2236047	MYCN	7	141898101	141898113	FOS::JUN
14	77805757	77805769	MYCN	14	2447896	2447908	FOS::JUN
14	77822455	77822467	MYCN	13	128590439	128590451	FOSL1
14	83370427	83370439	MYCN	13	139106535	139106547	FOSL1
14	83370516	83370528	MYCN	13	139055224	139055236	FOSL1
15	12909634	12909646	MYCN	13	139052320	139052332	FOSL1
16	34670999	34671014	Bach1::Mafk	13	139038832	139038844	FOSL1
13	139106534	139106549	Bach1::Mafk	4	68736661	68736673	FOSL1

13	139061316	13906132 6	NEUROD1	7	105745166	105745178	FOSL1
13	139051995	13905200 5	NEUROD1	14	77768272	77768284	FOSL1
7	141866168	14186617 8	NEUROD1	14	77817484	77817496	FOSL1
7	141877761	14187777 1	NEUROD1	13	139106535	139106547	FOSL2
7	141881555	14188156 5	NEUROD1	13	139055224	139055236	FOSL2
13	139106536	13910654 8	NFE2	13	139038833	139038845	FOSL2
13	139055224	13905523 6	NFE2	4	68508861	68508873	FOSL2
7	105714029	10571404 1	NFE2	4	68736660	68736672	FOSL2
7	105754009	10575402 1	NFE2	7	105750136	105750148	FOSL2
7	141898102	14189811 4	NFE2	7	141897741	141897753	FOSL2
13	139106537	13910654 9	Nfe2l2	7	141898101	141898113	FOSL2
13	139055225	13905523 7	Nfe2l2	14	2447896	2447908	FOSL2
7	105714030	10571404 2	Nfe2l2	16	34673281	34673293	FOXA1
7	141898103	14189811 5	Nfe2l2	4	68493391	68493403	FOXA1
16	35634651	35634661	NFIA	4	68497228	68497240	FOXA1
16	35634652	35634662	NFIC	4	68726830	68726842	FOXA1
13	128595659	12859566 9	NFIC	4	68731378	68731390	FOXA1
13	128064897	12806490 7	NFIC	4	68734240	68734252	FOXA1
13	139098576	13909858 6	NFIC	4	68735190	68735202	FOXA1
13	139062872	13906288 2	NFIC	4	68737280	68737292	FOXA1
13	139044583	13904459 3	NFIC	4	68741120	68741132	FOXA1
4	68735264	68735274	NFIC	7	105414018	105414030	FOXA1
4	68737296	68737306	NFIC	7	105414150	105414162	FOXA1
4	68746805	68746815	NFIC	7	105414185	105414197	FOXA1
7	141881588	14188159 8	NFIC	7	105421159	105421171	FOXA1
7	141898272	14189828 2	NFIC	7	105711577	105711589	FOXA1
14	29952103	29952113	NFIC	7	105749953	105749965	FOXA1
16	34694734	34694746	NFKB1	7	141867054	141867066	FOXA1
7	141874310	14187432 2	NFKB1	7	141903637	141903649	FOXA1
7	141883650	14188366 2	NFKB1	14	29950456	29950468	FOXA1
7	141896401	14189641 3	NFKB1	14	77759716	77759728	FOXA1
7	141903413	14190342 5	NFKB1	14	77760617	77760629	FOXA1
7	105784831	10578484 3	NFKB1	14	77763672	77763684	FOXA1
16	34694733	34694743	NFKB2	14	77782899	77782911	FOXA1
7	141866058	14186606 8	NFKB2	14	77791027	77791039	FOXA1
7	141874312	14187432 2	NFKB2	14	77791250	77791262	FOXA1
7	141903415	14190342 5	NFKB2	14	77791283	77791295	FOXA1
13	139059423	13905943 5	NFYA	14	77791443	77791455	FOXA1

13	139035350	13903536 2	NFYA	14	77802294	77802306	FOXA1
13	139059419	13905943 1	NFYB	14	77816872	77816884	FOXA1
13	139035346	13903535 8	NFYB	2	8422792	8422804	FOXA1
7	141925935	14192594 7	NFYB	2	8423667	8423679	FOXA1
12	83380643	83380651	Nkx3-1	2	8431219	8431231	FOXA1
7	141928460	14192846 8	BATF3	2	8434703	8434715	FOXA1
13	139055296	13905530 8	NR1H2::RXRA	12	83327075	83327087	FOXA1
13	139051473	13905148 5	NR1H2::RXRA	12	83380566	83380578	FOXA1
7	141892632	14189264 4	NR1H2::RXRA	16	35633911	35633923	Foxa2
7	141896490	14189650 2	NR1H2::RXRA	16	34694659	34694671	Foxa2
14	83370149	83370161	NR1H2::RXRA	7	105744161	105744173	Foxa2
13	128596210	12859622 2	NR2F1	4	68493388	68493400	Foxa2
13	128593951	12859396 3	NR2F1	4	68501677	68501689	Foxa2
13	128592643	12859265 5	NR2F1	4	68731375	68731387	Foxa2
13	139058994	13905900 6	NR2F1	4	68735187	68735199	Foxa2
4	68500587	68500599	NR2F1	4	68741123	68741135	Foxa2
7	141865464	14186547 6	NR2F1	4	68747527	68747539	Foxa2
7	141897936	14189794 8	NR2F1	7	105414188	105414200	Foxa2
14	2450157	2450169	NR2F1	7	105699313	105699325	Foxa2
14	77781150	77781162	NR2F1	7	105899842	105899854	Foxa2
14	77817419	77817431	NR2F1	7	141867051	141867063	Foxa2
7	141874529	14187454 1	NR3C1	7	141928187	141928199	Foxa2
13	139064509	13906452 1	NR3C2	14	29950453	29950465	Foxa2
7	141866187	14186619 9	NR3C2	14	77760614	77760626	Foxa2
13	128590438	12859045 0	BATF::JUN	14	77791280	77791292	Foxa2
13	128064754	12806476 6	BATF::JUN	15	12924989	12925001	Foxa2
13	139106537	13910654 9	BATF::JUN	2	8422795	8422807	Foxa2
13	139055225	13905523 7	BATF::JUN	12	83384719	83384731	Foxa2
13	139038834	13903884 6	BATF::JUN	13	128590825	128590833	ARNT::HIF1A
4	68508862	68508874	BATF::JUN	13	128065322	128065330	ARNT::HIF1A
4	68749615	68749627	BATF::JUN	16	34670999	34671009	FOXH1
7	141884682	14188469 4	BATF::JUN	14	29942334	29942342	Foxj2
7	141898100	14189811 2	BATF::JUN	16	34945138	34945148	Foxo1
14	2447898	2447910	BATF::JUN	14	77816926	77816936	Foxo1
13	128059518	12805952 8	OTX2	7	141884733	141884745	FOXP3
13	139035985	13903599 5	OTX2	16	34966012	34966022	Gabpa
13	139035121	13903513 1	OTX2	13	128592787	128592797	Gabpa
7	141866125	14186613 5	OTX2	13	128064919	128064929	Gabpa

4	68508864	68508874	PAX5	13	128064874	128064884	Gabpa
7	141865977	141865987	PAX5	13	128059477	128059487	Gabpa
13	128064757	128064767	PBX1	13	139053329	139053339	Gabpa
13	139098619	139098629	PBX1	13	139051443	139051453	Gabpa
13	139055224	139055234	PBX1	7	141865348	141865358	Gabpa
4	68732331	68732341	PBX1	7	141865876	141865886	Gabpa
7	105899043	105899053	PBX1	7	141877978	141877988	Gabpa
7	141896465	141896475	PBX1	7	141878158	141878168	Gabpa
7	141897742	141897752	PBX1	7	141884977	141884987	Gabpa
7	141898401	141898411	PBX1	7	141898108	141898118	Gabpa
13	128064690	128064698	PDX1	7	141928469	141928479	Gabpa
13	128594859	128594871	Bcl6	15	12907977	12907987	Gabpa
13	128592776	128592788	Bcl6	16	35651004	35651014	Gata1
14	29945470	29945482	Bcl6	16	34694610	34694620	Gata1
7	141874418	141874428	POU2F2	13	128067859	128067869	Gata1
7	141928178	141928188	POU2F2	13	139076062	139076072	Gata1
14	77791237	77791247	POU2F2	4	68732886	68732896	Gata1
16	34671862	34671872	POU5F1B	7	105704770	105704780	Gata1
13	139057837	139057847	POU5F1B	7	105899251	105899261	Gata1
7	105913656	105913666	POU5F1B	7	141868628	141868638	Gata1
14	29952440	29952450	POU5F1B	7	141878018	141878028	Gata1
13	139057838	139057850	Pou5f1::Sox2	7	141897217	141897227	Gata1
7	105706839	105706851	Pou5f1::Sox2	7	141897769	141897779	Gata1
7	141874416	141874428	Pou5f1::Sox2	7	141928414	141928424	Gata1
14	77821160	77821172	Pou5f1::Sox2	7	141928446	141928456	Gata1
16	34941865	34941877	Pparg::Rxra	14	29941653	29941663	Gata1
13	139051472	139051484	Pparg::Rxra	13	139053613	139053631	GATA1::TAL1
13	139038886	139038898	Pparg::Rxra	4	68508907	68508925	GATA1::TAL1
7	141865301	141865313	Pparg::Rxra	4	68739885	68739903	GATA1::TAL1
14	83370148	83370160	Pparg::Rxra	7	105902209	105902227	GATA1::TAL1
13	128594860	128594872	BCL6B	7	141899995	141900013	GATA1::TAL1
13	128592775	128592787	BCL6B	14	2445730	2445748	GATA1::TAL1
14	29945469	29945481	BCL6B	13	128067859	128067869	GATA2
7	141865976	141865984	Rarg	4	68497037	68497047	GATA2
16	34942479	34942489	REL	4	68732886	68732896	GATA2
13	128596580	128596590	REL	7	105782534	105782544	GATA2
7	105412026	105412036	REL	7	105889026	105889036	GATA2
7	105777892	105777902	REL	7	105899251	105899261	GATA2
7	141866058	141866068	REL	7	141868628	141868638	GATA2

7	141874312	14187432 2	REL	7	141878018	141878028	GATA2
7	141896403	14189641 3	REL	7	141897217	141897227	GATA2
7	141903415	14190342 5	REL	7	141928446	141928456	GATA2
16	34942479	34942489	RELA	14	29941653	29941663	GATA2
7	105777874	10577788 4	RELA	12	83320206	83320216	GATA2
7	141866058	14186606 8	RELA	16	35653044	35653056	GATA3
7	141896403	14189641 3	RELA	13	139061686	139061698	GATA3
7	141903415	14190342 5	RELA	13	139054176	139054188	GATA3
13	139054979	13905499 1	REST	4	68497196	68497208	GATA3
14	83381857	83381869	BHLHE40	7	105899250	105899262	GATA3
14	83370427	83370439	BHLHE40	7	105902257	105902269	GATA3
16	34978783	34978795	RUNX1	7	141865490	141865502	GATA3
13	128064877	12806488 9	RUNX1	7	141878017	141878029	GATA3
13	139073721	13907373 3	RUNX1	7	141905798	141905810	GATA3
13	139051424	13905143 6	RUNX1	7	141928445	141928457	GATA3
4	68508509	68508521	RUNX1	14	29941652	29941664	GATA3
4	68736120	68736132	RUNX1	14	77763648	77763660	GATA3
7	141884010	14188402 2	RUNX1	14	77782475	77782487	GATA3
7	141884947	14188495 9	RUNX1	2	8525661	8525673	GATA3
7	141903522	14190353 4	RUNX1	12	83327109	83327121	GATA3
7	141925375	14192538 7	RUNX1	12	83357336	83357348	GATA3
7	141928334	14192834 6	RUNX1	16	35651001	35651013	Gata4
7	141928935	14192894 7	RUNX1	13	128067860	128067872	Gata4
14	29945936	29945948	RUNX1	4	68497038	68497050	Gata4
14	2236407	2236419	RUNX1	7	105882644	105882656	Gata4
15	12909602	12909614	RUNX1	7	141868550	141868562	Gata4
13	128064878	12806488 8	RUNX2	7	141868629	141868641	Gata4
16	34978784	34978794	RUNX3	7	141878019	141878031	Gata4
13	128590463	12859047 3	RUNX3	7	141928443	141928455	Gata4
13	128064878	12806488 8	RUNX3	14	29941650	29941662	Gata4
13	139073722	13907373 2	RUNX3	14	2237390	2237402	Gata4
4	68508510	68508520	RUNX3	15	12905363	12905375	Gata4
7	141877982	14187799 2	RUNX3	2	8414369	8414381	Gata4
7	141884011	14188402 1	RUNX3	12	83320203	83320215	Gata4
7	141925376	14192538 6	RUNX3	12	83357334	83357346	Gata4
7	141928936	14192894 6	RUNX3	12	83357385	83357397	Gata4
14	29954760	29954770	RUNX3	12	83382436	83382448	Gata4
16	34941864	34941876	Rxra	7	141868548	141868560	GATA5
13	139059770	13905978 2	Rxra	7	141878017	141878029	GATA5

13	139057676	13905768 8	Rxra	7	141928445	141928457	GATA5
13	139051473	13905148 5	Rxra	14	29941652	29941664	GATA5
13	139038887	13903889 9	Rxra	12	83322177	83322189	GATA5
7	141865302	14186531 4	Rxra	13	139063135	139063145	Gfi1b
7	141892632	14189264 4	Rxra	16	34967117	34967127	Hand1::Tcf3
7	141896490	14189650 2	Rxra	13	128595659	128595669	Hand1::Tcf3
14	77770481	77770493	Rxra				
16	34979435	34979447	BHLHE41				
13	139059342	13905935 4	BHLHE41				
14	77814427	77814439	BHLHE41				
14	83370351	83370363	BHLHE41				
12	83382850	83382862	BHLHE41				
16	35634684	35634696	Sox2				
13	128059231	12805924 3	Sox2				
13	139057490	13905750 2	Sox2				
7	105703378	10570339 0	Sox2				
7	105888384	10588839 6	Sox2				
7	141865302	14186531 4	Sox2				
7	141874539	14187455 1	Sox2				
7	141896557	14189656 9	Sox2				
14	29949960	29949972	Sox2				
14	2445200	2445212	Sox2				
14	77778981	77778993	Sox2				
14	77819815	77819827	Sox2				
14	77821184	77821196	Sox2				
15	12909686	12909698	Sox2				
2	8447253	8447265	Sox2				
12	83316661	83316673	Sox2				
13	139035349	13903535 9	SOX9				
4	68498861	68498871	SOX9				

M5. MITOKONDRIÁLIS REKOMBINÁCIÓ – SCRIPTEK

1. mtrec_detector.py – main script

```

1. #!/usr/bin/python3
2.
3. import sys, argparse, os
4. import collections
5. import bamnostic as bs
6. import seqhandlers as sq
7. import ioprocessing as io
8.
9. parser = argparse.ArgumentParser()

```

```

10.
11. #Mandatory args
12. parser.add_argument('-b', '--bamfile', help = "input bam file")
13. parser.add_argument('-v', '--vcffile', help = "input vcf file")
14. parser.add_argument('-p', '--out_positive',
15.                     help = 'filename for positive hit output')
16.
17. #Optional
18. parser.add_argument('--out_negative',
19.                     action = 'store_true',
20.                     help = 'output negative results, too')
21. parser.add_argument('--minstretch', type = int, default = 2,
22.                     help = 'Minimum length of type X to score positively')
23. parser.add_argument('--opposite', nargs = '*', type = str,
24.                     default = ['14189', '14366'],
25.                     help = 'SNPs called, but more prominent in default type')
26. parser.add_argument('--minscore', type = int, default = 2,
27.                     help = 'Minimum score to consider a read a positive hit')
28.
29. parser.add_argument('--plotdata', action = 'store_true',
30.                     help = 'Output per site data for plotting')
31.
32. args = parser.parse_args()
33.
34. print('Reading data...')
35. #cigar tags to exclude to get reference coordinates
36. ref_exclude = [4, 1, 5, 6, 8]
37.
38. #Get data from files
39. bamfile = bs.AlignmentFile(args.bamfile, "rb")
40. read_info = [ (read.query_name,
41.               read.reference_start,
42.               read.reference_end,
43.               #read.query_alignment_sequence,
44.               read.query_alignment_length,
45.               read.cigar)
46.               for read in bamfile ]
47.
48. var = sq.get_variants(args.vcffile)
49.
50. print('Input files:')
51. print(f'\t{args.bamfile}')
52. print(f'\t{args.vcffile}')
53. print(f'Output file:\n\t {args.out_positive}')
54. print(f'Switched SNPs: {args.opposite}')
55. print(f'Minimum length of type to score positively: {args.minstretch}')
56. print(f'Outputting negative results: {args.out_negative}')
57.
58. print('Processing...')
59.
60.
61. fullreaddic = io.create_full_dic(read_info, var, args.opposite)
62. filtdic = io.filter_readdic(fullreaddic, args.minstretch, args.minscore)
63. delta = sq.supporting_reads(read_info)
64. #print(delta)
65.
66. print('Writing output...')
67. with open(args.out_positive, 'w') as out:
68.     out.write(f'read\tcigar\tresult\ttscore\ttype\n')
69.     for read, info in filtdic.items():
70.         if read in delta:
71.             out.write(read+'\t'+'.join(info[1])+'\t'+'\t'.join(info[2])+'\tdelta\n')
72.         else:
73.             out.write(read+'\t'+'.join(info[1])+'\t'+'\t'.join(info[2])+'\ttw\n')
74.
75. if args.out_negative:
76.     with open('negative.out', 'w') as neg:

```

```

77.     neg.write(f'read\tcigar\tresult\tscore\ttype\n')
78.     for read, readdic in fullreaddic.items():
79.         ps_cigar = sq.read_summary(list(readdic.values()))
80.         res = sq.rec_possible(ps_cigar, args.minstretch, args.minscore)
81.
82.         if res[0] == 'False':
83.             if read in delta:
84.                 neg.write(f'{read}\t' + ''.join(ps_cigar) + '\t' + '\t'.join(res) +
'\tdelta\n')
85.             else:
86.                 neg.write(f'{read}\t' + ''.join(ps_cigar) + '\t' + '\t'.join(res) +
'\tw\n')
87.
88. if args.plotdata:
89.     dat = io.plot_data(fullreaddic)
90.     with open('plotdata.tsv', 'w') as pld:
91.         for line in dat:
92.             pld.write("\t".join(str(i) for i in line))
93.             if line[0] in delta:
94.                 pld.write('\tdelta\n')
95.             else:
96.                 pld.write('\tw\n')
97.

```

2.ioprocessing.py – Input/output format processing functions

```

1. import seqhandlers as sq
2. import bamnostic as bs
3. import collections
4.
5. #cigar tags to exclude to get reference coordinates
6. ref_exclude = [4, 1, 5, 6, 8]
7.
8. #Summarizing the data
9. def create_full_dic(read_info, var, opposite):
10.     """Put together info of all reads"""
11.
12.     readnames = list(set([read[0] for read in read_info]))
13.     fullreaddic = {readname : [] for readname in readnames}
14.
15.     for read in read_info:
16.         ref_snps = sq.snp_coordinates(read, ref_exclude)
17.         nonsnps = sq.nonsnp_coordinates(var, read, ref_snps)
18.
19.         readdic = sq.opposite_snps(opposite,
20.                                   sq.create_readdic(var, read, ref_snps, nonsnps))
21.
22.         if len(fullreaddic[read[0]]) != 0:
23.             fullreaddic[read[0]] = fullreaddic[read[0]] | readdic
24.         else:
25.             fullreaddic[read[0]] = readdic
26.
27.     return(fullreaddic)
28.
29. #Filtering
30. def filter_readdic(fullreaddic, minstretch, minscore):
31.     """Filter interesting reads from all"""
32.     filtereddic = collections.defaultdict(lambda : [])
33.
34.     for read, readdic in fullreaddic.items():
35.         ps_cigar = sq.read_summary(list(readdic.values()))
36.         res = sq.rec_possible(ps_cigar, minstretch, minscore)
37.
38.         if res[0] == 'True':
39.             filtereddic[read] = [readdic, ps_cigar, res]
40.

```

```

41.     return(filterediddic)
42.
43. #Data for plotting
44. def plot_data(fullreaddic):
45.     """Create tsv that makes plotting easy"""
46.     data = [[k, a, b]
47.             for k, v in fullreaddic.items()
48.             for a, b in v.items()]
49.
50.     return(data)
51.
52. #Output
53. def dicprint(filtddic):
54.     f.open()
55.     out.write(f'{read}\t' + ''.join(ps_cigar) + '\t' + '\t'.join(res) + '\n')
56.

```

3.seqhandlers.py – Sequence and data processing and calculations

```

1. import collections
2. import bamnostic as bs
3.
4. #Data
5. #Generic, works for every sam file
6. #cigar tags to exclude to get reference coordinates
7. ref_exclude = [4, 1, 5, 6, 8]
8. #cigar tags to exclude to get query coordinates
9. q_exclude = [2, 3, 4, 5, 6, 8]
10.
11. #Functions
12.
13. #Chr, pos, ref, alt from vcf
14. def get_variants(vcf_file):
15.     """Read in variant coordinates from vcf"""
16.
17.     with open(vcf_file) as vcf:
18.         #Get chr, id, pos, ref, alt
19.         variant_info = [line.strip().split('\t')[0:5]
20.                         for line in vcf
21.                         if not line.startswith('#')]
22.
23.         #Have to add indel-control, or just filter them from vcf
24.         variant_info = [var for var in variant_info
25.                         if len(var[3]) == 1]
26.
27.     return(variant_info)
28.
29.
30. #Figure out the variants
31. def snp_coordinates(read, exclusion):
32.     """Function to get sequence coordinates"""
33.     """Use predefined cigar tag list as exclusion"""
34.
35.     rq_snps = []
36.     current_coord = read[1] #start
37.
38.     for op in read[4]: #cigar
39.         if op[0] not in exclusion:
40.             current_coord += int(op[1])
41.         elif op[0] == 8:
42.             for i in range(0,op[1]):
43.                 current_coord += 1
44.                 rq_snps.append(current_coord)
45.                 i += 1
46.
47.     return(rq_snps)

```

```

48.
49. def nonsnp_coordinates(var, read, rq_snps):
50.     """Get variant positions with no actual variant"""
51.     non_snps = [va[1] for va in var
52.                 if int(va[1]) < read[2] and
53.                 int(va[1]) > read[1]
54.                 and int(va[1]) not in rq_snps]
55.
56.     non_snps = non_snps
57.     return(non_snps)
58.
59. def create_readdic(var, read, ref_snps, non_snps):
60.     """Generic sequence creator"""
61.     readseq = ""
62.     readcoords = ""
63.
64.     readseq = [readseq + va[3] for va in var
65.               if int(va[1]) > read[1] #start
66.               and int(va[1]) < read[2]] #end
67.     readcoords = [readcoords + va[1] for va in var
68.                  if int(va[1]) > read[1]
69.                  and int(va[1]) < read[2]]
70.     read_dic = dict(zip(readcoords, readseq))
71.
72.     for pos, va in read_dic.items(): #check if this works as intended
73.         if pos in non_snps:
74.             read_dic[pos] = "A"
75.         elif int(pos) in ref_snps:
76.             read_dic[pos] = "B"
77.
78.     return(read_dic)
79.
80. #Pseudocigar for pseudoseq
81. def read_summary(pseudoseq):
82.     i = 0
83.     result = []
84.
85.     pseudoseq.append(object()) # append a dummy item first
86.     while(i < len(pseudoseq)-1):
87.         count = 1
88.         result.append(pseudoseq[i])
89.         while(pseudoseq[i] == pseudoseq[i+1]):
90.             count += 1
91.             i += 1
92.
93.         result.append(str(count))
94.         i += 1
95.
96.     pseudoseq.pop()
97.     return(result)
98.
99. #Very basic check for recombination possibility
100. def rec_possible(ps_cigar, minstretch, minscore):
101.     """Scoring the possibility of recombination"""
102.
103.     score = 0
104.     if len(ps_cigar) > 2:
105.         for i in ps_cigar[1::2]:
106.             #print(i)
107.             if int(i) >= minstretch:
108.                 score += 1
109.             elif int(i) == 1:
110.                 score -= 1
111.     if score >= minscore:
112.         return(["True", str(score)])
113.     else:
114.         return(["False", str(score)])

```

```

115.
116.
117. def opposite_snps(opposite, readdic):
118.     """Called snps that are actually in the ref type"""
119.     for pos, base in readdic.items():
120.         if pos in opposite:
121.             if readdic[pos] == "A":
122.                 readdic[pos]= "B"
123.             else:
124.                 readdic[pos] = "A"
125.
126.     return(readdic)
127.
128. #Check for deletion type
129. def supporting_reads(read_info):
130.     """Get delta info"""
131.
132.     delstart = 10522
133.     delend = 16287
134.     #Get all reads contained in the circ region
135.
136.     suppreads = []
137.     #if read ends at delstart or starts at delend
138.     for read in read_info:
139.         if read[2] == 10522 or read[1] == 15487:
140.             if read[0] not in suppreads:
141.                 suppreads.append(read[0])
142.
143.     return(suppreads)
144.

```

4.Kiegészítő scriptek

4.1. *select_pure_mt.py* – Tisztán mitokondriális readek kiválasztása az illesztésből

```

1. #!/usr/bin/python3
2.
3. import sys
4. import os
5. import argparse
6. import subprocess
7. import collections
8. import bamnostic as bs
9.
10. parser = argparse.ArgumentParser()
11.
12. parser.add_argument("-i", "--infile", help = 'input bam file')
13. parser.add_argument('-o', '--outfile', help = 'textfile listing pure MT readnames',
14.                     default = 'out.txt')
15. parser.add_argument('-b', '--bamout', help = 'pure MT read bam output',
16.                     default = 'puremt.bam')
17. args = parser.parse_args()
18.
19. #Read input bam
20. bam = bs.AlignmentFile(args.infile, 'rb')
21.
22. #Create a list of not pure MT reads
23. def get_nonpure_ids(bam):
24.     """Find reads with supplementary alignments not on MT"""
25.     nonmt_list = []
26.     for read in bam:
27.         try:
28.             sa_tag = read.get_tag('SA')
29.             rname = read.query_name
30.

```

```

31.         for tag in sa_tag.strip().split(';'):
32.             if not tag.startswith('MT') and len(tag) > 2:
33.                 nonmt_list.append(read.query_name)
34.                 print(tag.split(',')[0])
35.
36.         except KeyError:
37.             pass
38.
39.     return(list(set(nonmt_list)))
40.
41. nonpure = get_nonpure_ids(bam)
42.
43. bam = bs.AlignmentFile(args.infile, 'rb')
44. readidlist = [read.query_name for read in bam
45.               if read.query_name not in nonpure]
46.
47. #Write pure MT read ids to file
48. with open(args.outfile, 'w') as out:
49.     for i in readidlist:
50.         out.write(f'{i}\n')
51.
52. #Run samtools to get pure MT reads in a bam file
53. command = f'samtools view -b -o {args.bamout} -N {args.outfile} {args.infile}'
54. subprocess.call(command, shell = True)
55.

```

4.2. Ábrák generálásához használt R scriptek

hp202_plot.R

```

1. library(tidyverse)
2. library(ggpubr)
3. library(ggtranscript)
4.
5. #Get and format data
6. plotdata <- read.table('../Results/HP202.plotdata.2.3.tsv', sep = '\t', header = F)
7. colnames(plotdata) <- c('Rname', 'Position', 'Genotype', 'DelType')
8.
9. #Subset by readnames for plotting
10. rname_df <- read.table('../Results/HP202.2.3.pos.rnames.txt', sep = '\t', header = F)
11. colnames(rname_df) <- c('Rname', 'Status')
12. rnames <- c(rname_df$Rname)
13. plotdata <- subset(plotdata, Rname %in% rnames)
14. plotdata <- left_join(plotdata, rname_df, by = "Rname")
15.
16.
17. #Calculate values for path plotting
18. startsite <- function(incoord, x) {
19.     startpos <- incoord[x] - ((incoord[x] - incoord[x -1])/2)
20.     return(startpos)
21. }
22.
23. plotdata$Start <- 0
24.
25. for (x in 1:nrow(plotdata)) {
26.     if (x > 1 && x < nrow(plotdata)) {
27.         if (plotdata$Rname[x] == plotdata$Rname[x-1]) {
28.             plotdata$Start[x] <- startsite(plotdata$Position, x)
29.         }
30.         else {
31.             plotdata$Start[x] <- plotdata$Position[x]
32.         }
33.     }
34.     else {
35.         plotdata$Start[x] <- plotdata$Position[x]

```

```

36.   }
37. }
38.
39. #replace last start value, for the end of the path plotting
40. plotdata$Start[plotdata$Position == 16276] <- 16276
41.
42.
43. #The plot
44. posdata <- plotdata[plotdata$Status == "True", ]
45.
46. p <- ggplot(posdata, aes(Position,Rname,group=Rname,col=Genotype)) +
47.   geom_point(size = 1.5) +
48.   geom_path(aes(Start,Rname,group=Rname,col=Genotype), linewidth = 1) +
49.   theme_minimal() +
50.   scale_color_manual(values=c("Orange", "Blue")) +
51.   theme(axis.text.y = element_blank(),
52.         panel.grid.minor.x = element_blank()) +
53.   scale_x_continuous(name = 'Position',
54.                     breaks = c(unique(plotdata$Position)),
55.                     labels = c(unique(plotdata$Position)),
56.                     guide = guide_axis(check.overlap = T),
57.                     limits = c(1, 16276)) +
58.   theme(axis.text.x = element_text(angle = 90)) +
59.   ylab("Reads") +
60.   theme(legend.position = "top")
61. p
62.
63. annot <- read.table('../Results/NZB.summary.tsv', sep = '\t', header = F)
64. annot$V10 <- (annot$V2 + annot$V3)/2
65.
66. a <- ggplot(annot) +
67.   geom_range(aes(xstart = V2, xend = V3, y = V1, fill = V6, )) +
68.   theme_minimal() +
69.   scale_x_continuous(name = 'Gene',
70.                     breaks = c(annot$V10),
71.                     labels = c(annot$V7),
72.                     #guide = guide_axis(check.overlap = T),
73.                     limits = c(1, 16276)) +
74.   theme(axis.text.x = element_text(angle = 90)) +
75.   guides(fill=guide_legend(title="Gene type")) +
76.   ylab("Annotation") +
77.   theme(axis.text.y = element_blank(),
78.         panel.grid.minor.x = element_blank(),
79.         legend.position = 'none')
80.
81. fig <- ggarrange(p, a,
82.                 labels = c("Recombinant reads", ""),
83.                 ncol = 1, nrow = 2,
84.                 heights = c(8, 1))
85. fig
86.
87. ggsave(file="HP202_positive.annotated.pdf", plot = p, width = 210, height = 297, units = "mm")
88.

```

delta_vis.R

```

1. library('tidyverse')
2. library('ggpubr')
3. library(ggtranscript)
4.
5.
6. #Get and format data
7. plotdata <- read.table('../Results/BH10.full.23-10-31.plotdata.tsv', sep = '\t', header = F)
8. colnames(plotdata) <- c('Rname', 'Position', 'Genotype', 'DelType')
9.

```

```

10. #Subset by readnames for plotting
11. rname_df <- read.table('reads.txt', sep = '\t', header = F)
12. colnames(rname_df) <- c('Rname', 'Status')
13. rnames <- c(rname_df$Rname)
14. plotdata <- subset(plotdata, Rname %in% rnames)
15. plotdata <- left_join(plotdata, rname_df, by = "Rname")
16.
17. #Get data for delta reads
18. deltanames <- subset(plotdata, DelType == 'delta')
19. deltanames <- unique(deltanames$Rname)
20.
21. deltaloc <- c(10522, 15487)
22. Rname <- rep(c(deltanames), each = 2)
23. loc <- rep(deltaloc, length(Rname))
24. deltable <- data.frame(Rname, loc)
25. deltable <- left_join(deltable, rname_df, by = "Rname")
26. #deltable$DelType <- 'delta'
27.
28. #Calculate values for path plotting
29. startsite <- function(incoord, x) {
30.   startpos <- incoord[x] - ((incoord[x] - incoord[x - 1])/2)
31.   return(startpos)
32. }
33.
34. plotdata$Start <- 0
35.
36. for (x in 1:nrow(plotdata)) {
37.   if (x > 1 && x < nrow(plotdata)) {
38.     if (plotdata$Rname[x] == plotdata$Rname[x-1]) {
39.       plotdata$Start[x] <- startsite(plotdata$Position, x)
40.     }
41.     else {
42.       plotdata$Start[x] <- plotdata$Position[x]
43.     }
44.   }
45.   else {
46.     plotdata$Start[x] <- plotdata$Position[x]
47.   }
48. }
49.
50. #replace last start value, for the end of the path plotting
51. plotdata$Start[plotdata$Position == 16207] <- 16207
52.
53.
54. #The plot
55. posdata <- plotdata[plotdata$Status == "True", ]
56. posdel <- deltable[deltatable$Status == "True", ]
57. negdata <- plotdata[plotdata$Status == "False", ]
58. negdel <- deltable[deltatable$Status == "False", ]
59.
60. p <- ggplot(posdata, aes(Position, Rname, group=Rname, col=Genotype)) +
61.   geom_point(size = 2.5) +
62.   geom_path(aes(Start, Rname, group=Rname, col=Genotype), linewidth = 1.5) +
63.   theme_minimal() +
64.   scale_color_manual(values=c("Orange", "Blue")) +
65.   geom_path(data = posdel, aes(loc, Rname, group = Rname),
66.     colour = 'lightgrey', linewidth = 1.5) +
67.   theme(axis.text.y = element_blank(),
68.     panel.grid.minor.x = element_blank()) +
69.   scale_x_continuous(name = 'Position',
70.     breaks = c(unique(plotdata$Position), deltaloc),
71.     labels = c(unique(plotdata$Position), deltaloc),
72.     guide = guide_axis(check.overlap = T),
73.     limits = c(1, 16207)) +
74.   theme(axis.text.x = element_text(angle = 90)) +
75.   ylab("Reads") +
76.   theme(legend.position = "top")

```

```

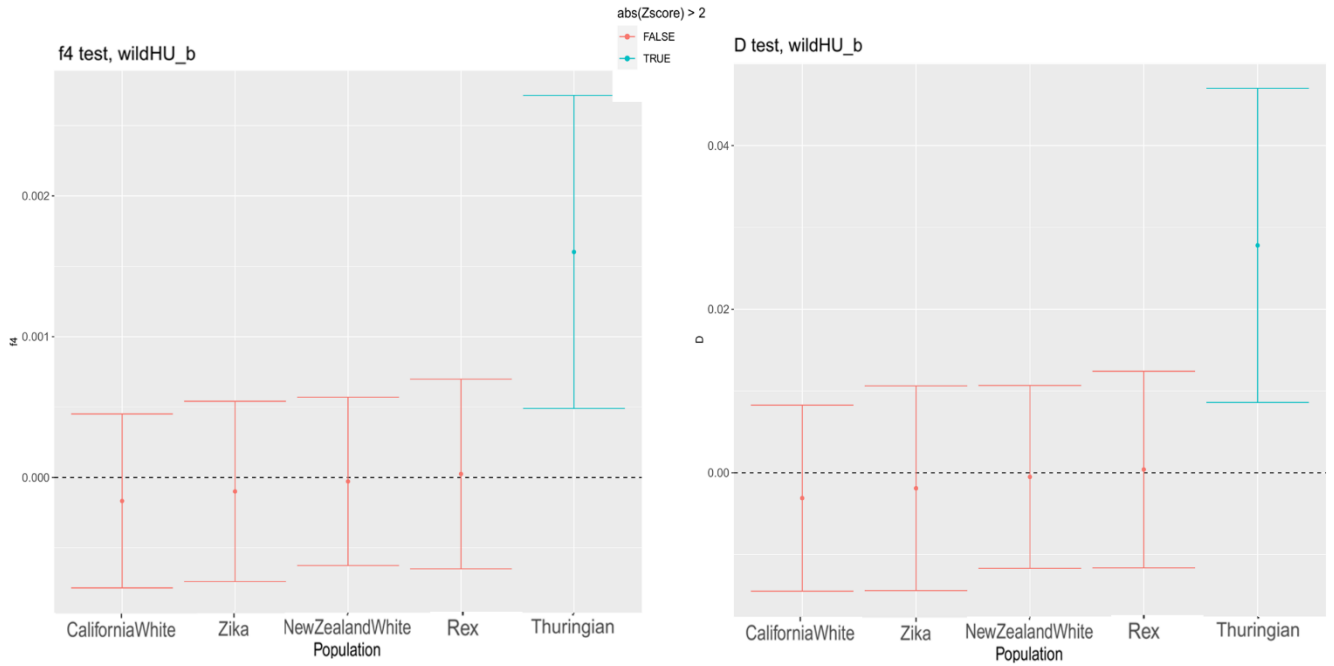
77.
78. n <- ggplot(negdata, aes(Position,Rname,group=Rname,col=Genotype)) +
79.   geom_point(size = 2.5) +
80.   geom_path(aes(Start,Rname,group=Rname,col=Genotype), linewidth = 1.5) +
81.   theme_minimal() +
82.   scale_color_manual(values=c("Orange", "Blue"),
83.                      guide = guide_legend(override.aes = list(color = "white"))) +
84.   geom_path(data = negdel, aes(loc, Rname, group = Rname),
85.            colour = 'lightgrey', linewidth = 1.5) +
86.   theme(axis.text.y = element_blank(),
87.         panel.grid.minor.x = element_blank()) +
88.   scale_x_continuous(name = 'Position',
89.                     breaks = c(unique(plotdata$Position), deltaloc),
90.                     labels = c(unique(plotdata$Position), deltaloc),
91.                     guide = guide_axis(check.overlap = T),
92.                     limits = c(1, 16207)) +
93.   theme(axis.text.x = element_text(angle = 90),
94.         legend.title = element_text(color = "transparent"),
95.         legend.text = element_text(color = "transparent")) +
96.   ylab("Reads") +
97.   theme(legend.position = "top")
98.
99.
100. annot <- read.table('./Results/BH10.summary.tsv', sep = '\t', header = F)
101. annot$V10 <- (annot$V2 + annot$V3)/2
102.
103. a <- ggplot(annot) +
104.   geom_range(aes(xstart = V2, xend = V3, y = V1, fill = V6, )) +
105.   theme_minimal() +
106.   scale_x_continuous(name = 'Gene',
107.                     breaks = c(annot$V10),
108.                     labels = c(annot$V7),
109.                     guide = guide_axis(check.overlap = T),
110.                     limits = c(1, 16207)) +
111.   theme(axis.text.x = element_text(angle = 90)) +
112.   guides(fill=guide_legend(title="Gene type")) +
113.   ylab("Annotation") +
114.   theme(axis.text.y = element_blank(),
115.         panel.grid.minor.x = element_blank(),
116.         legend.position = 'none')
117.
118.
119.
120.
121.
122. fig <- ggarrange(p, n,
123.                 labels = c("Recombinant", "Non-recombinant"),
124.                 ncol = 1, nrow = 2)
125. #fig
126. f2 <- ggarrange(fig, a,
127.                 ncol = 1, nrow = 2,
128.                 heights = c(3, 1))
129. f2
130.

```

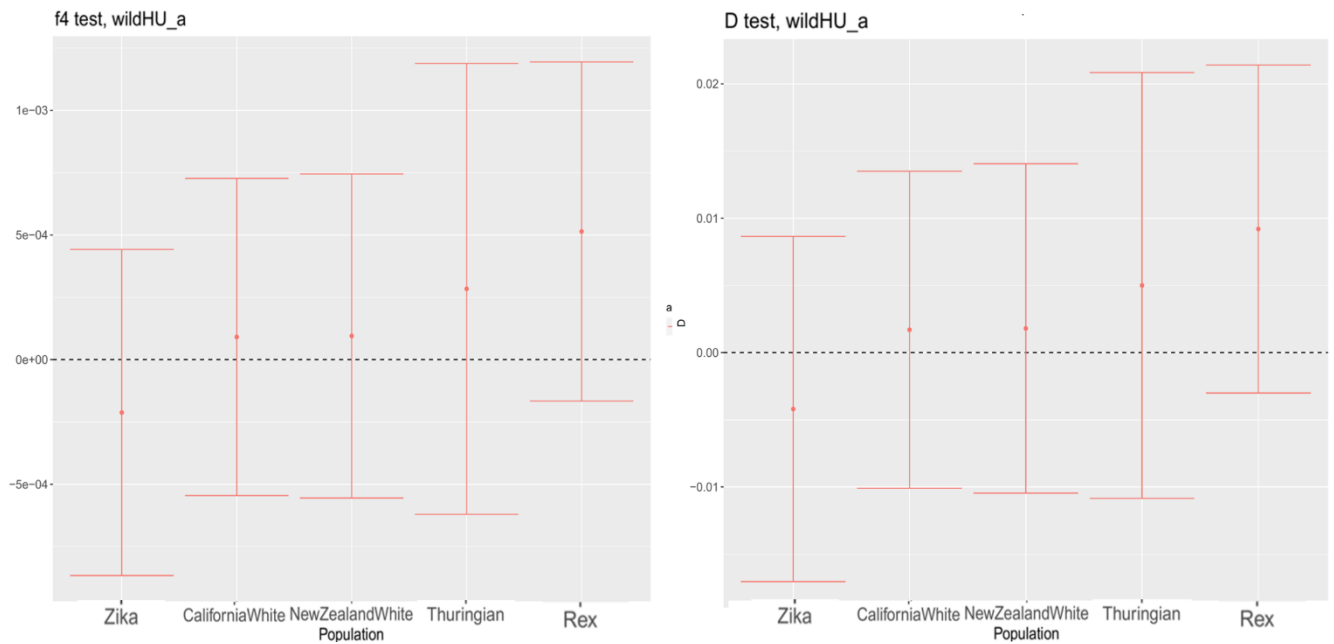
M6. D ÉS F4 TESZTEK EREDMÉNYEI – ÜREGI- ÉS HÁZINYÚL HIBRIDIZÁCIÓ

F4 (balra) és D tesztek (jobbra) eredményei az üregi- és házinyúl populációk között

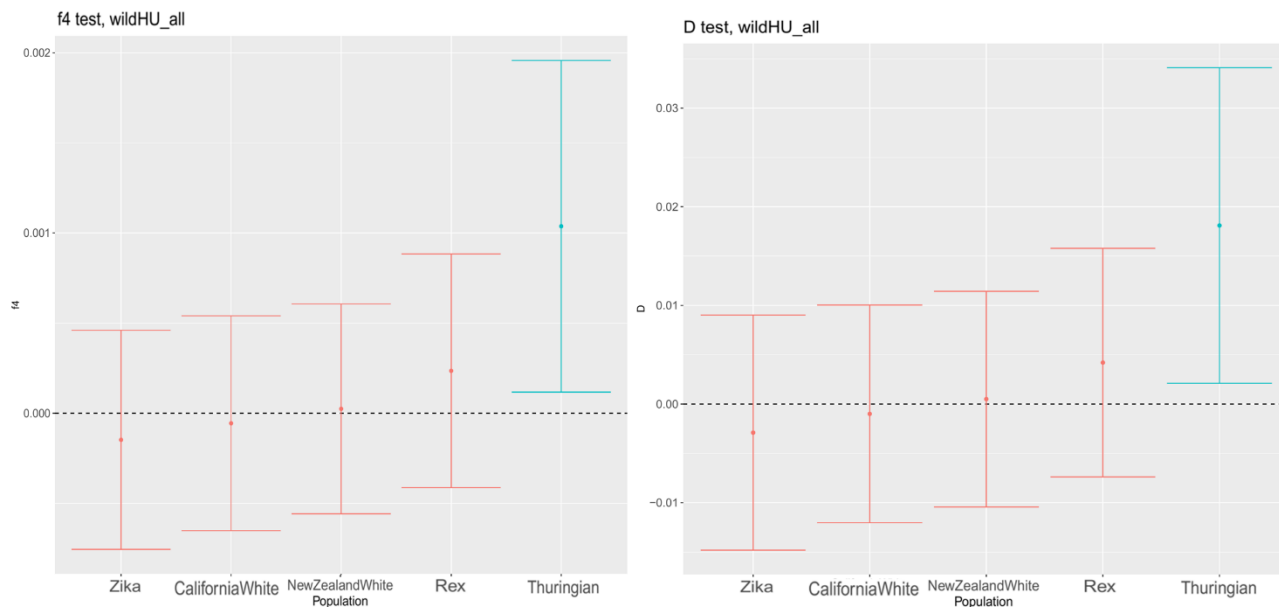
A) *Tesztek a wildHU_b (Budapesti Állat- és Növénykert) üregi nyúl populációra*



B) *Tesztek a wildHU_a (Tiszaszentimre) üregi nyúl populációra*



C) Összevont tesztek az összes magyarországi üregi nyúl mintával



M7. LEPUS TIMIDUS GENOM-ÖSSZESZERELÉS – STATISZTIKÁK

1. Az összeszerelt genom számszerű jellemzői (Fekete *et al.*, 2025a)

Assembly statistics	Primary	Alternate
# scaffolds	85	109
Total scaffold length	2,695,305,354	2,556,802,277
Average scaffold length	31,709,474.75	23,456,901.62
Scaffold N50	125,755,317	126,936,470
Scaffold auN	125,197,910.08	126,023,208.11
Scaffold L50	9	9
Largest scaffold	182,608,553	185,135,719
Smallest scaffold	10,771	21,369
# contigs	1,414	1,454
Total contig length	2,695,169,802	2,556,665,250
Average contig length	1,906,060.68	1,758,366.75
Contig N50	4,885,811	5,114,768
Contig auN	6,632,874.02	6,902,786.36
Contig L50	157	136
Largest contig	32,947,725	26,587,001
Smallest contig	967	21,369
# gaps in scaffolds	1,329	1,345
Total gap length in scaffolds	135,552	137,027
Average gap length in scaffolds	102	101.88
Gap N50 in scaffolds	100	100
Gap auN in scaffolds	104.1	103.78
Gap L50 in scaffolds	650	660
Largest gap in scaffolds	200	200
Smallest gap in scaffolds	26	27
Base composition (A:C:G:T)		
A	758,096,831	716,813,053
C	589,442,155	561,406,584
G	589,603,553	561,320,925
T	758,027,263	717,124,688
GC content %	43.75	43.91
# soft-masked bases	0	0
# segments	1,414	1,454
Total segment length	2,695,169,802	2,556,665,250
Average segment length	1,906,060.68	1,758,366.75
# gaps	1,329	1,345
# paths	85	109

2.BUSCO eredmények (Fekete *et al.*, 2025a)

BUSCO	Primary		Alternate	
	Count	%	Count	%
Mammalia				
Complete BUSCOs (C)	8,770	95.06	8,601	93.23
Complete and single-copy BUSCOs (S)	8,517	92.32	8,378	90.81
Complete and duplicated BUSCOs (D)	253	2.74	223	2.42
Fragmented BUSCOs (F)	76	0.82	84	0.91
Missing BUSCOs (M)	380	4.12	541	5.86
Total BUSCO groups searched	9,226		9,226	
Glires				
Complete BUSCOs (C)	12,863	93.22	12,553	90.98
Complete and single-copy BUSCOs (S)	12,495	90.56	12,226	88.61
Complete and duplicated BUSCOs (D)	368	2.67	327	2.37
Fragmented BUSCOs (F)	157	1.14	149	1.08
Missing BUSCOs (M)	778	5.64	1096	7.94
Total BUSCO groups searched	13,798		13,798	

KÖSZÖNETNYILVÁNÍTÁS

Köszönetemet fejezem ki témavezetőimnek Dr. Kontra Leventének és Dr. Hoffmann Orsolya Ivettnek munkámban nyújtott segítségükért, szakmai és emberi támogatásukért,

A MATE Genetika- és Biotechnológia Intézet, Bioinformatika Csoportjának,

A University of Eastern Finland, Department of Environmental and Biological Sciences (Itä-suomen yliopisto, Ympäristö- ja biotieteiden laitos) Mitokondriális biológia csoportjának,

Valamint családomnak és barátaimnak,

akik lehetővé tették és támogatták munkámat.