

DOKTORI (PhD) ÉRTEKEZÉS TÉZISEI

**SZENT ISTVÁN EGYETEM– KAPOSVÁRI CAMPUS
AGRÁR ÉS KÖRNYEZETTUDOMÁNYI KAR**

Állattenyésztés-technológia és Menedzsment Tanszék

A doktori iskola vezetője:
PROF. DR. SZABÓ ANDRÁS
MTA doktora

Témavezetők:
PROF. DR. HORN PÉTER
MTA rendes tagja

PROF. DR. OROSZ LÁSZLÓ
MTA rendes tagja

**A GÍMSZARVAS KROMOSZÓMÁK DNS SZEKVENCIÁJA, A
CERELA1.0 GÍMSZARVAS GENOM ÖSSZEÁLLÍTÁSA**

Készítette:
BANA ÁGNES NÓRA

**KAPOSVÁR
2020**

1. A KUTATÁS ELŐZMÉNYEI

A gímszarvas (*Cervus elaphus hippelaphus*) ősidők óta szerves részét képezi az emberi kultúrának, kiemelkedő társadalmi, természeti és gazdasági haszonnal rendelkezik. Holarktikus elterjedésű faj, amelynek egyedei nagyszámban megtalálhatók Eurázsia, Észak-Amerika és bizonyos észak-afrikai területek erdőségeiben és pusztáin. Világszerte elterjedtek a szarvasfarmok, ahol modern tenyésztési eljárásokkal jó minőségű húskészítményeket állítanak elő és ahonnan magas CIC pontszámú trófeák származnak. A magyarországi fauna rekord méretű trófeáival (Szálka, Gamás, Lenti, Pusztakovácsi, Gemenc-Karapanca, Vajszló, Lábod) emelkedik ki az európai populációk közül.

A gemenci gímszarvas állományhoz köthető a bőszenfai Szarvasfarm, amely a Kaposvári Egyetem és a Bőszenfai Vadgazdálkodási Tájéközpont együttműködéséből született meg 1991-92 között. Az 1998-ban induló gímszarvas genom program szorosan kötődik a Bőszenfai Szarvasfarmhoz és a Kaposvári Egyetem Állattenyésztési Tudományok Doktori Iskolájához. A projekt megvalósításában a gödöllői Mezőgazdasági Biotechnológiai Központ és a budapesti ELTE Genetikai Tanszék és a SOTE 1. Belgyógyászati Klinika MSc és PhD programjai is jelentős részt vállaltak.

A CerEla1.0 alapjául szolgáló DNS minta a Kaposvári Egyetem Vadgazdálkodási Tájéközpont Bőszenfai Szarvasfarmján, egy természet közeli körülmények között élő 7 éves, kapitális gímszarvasbikából (fülszáma: Crot. N.o. 3016) származott. A 3x10 ml vér levételét követően történt meg a teljes genomi DNS kivonása Duplica Prep Automated DNA/RNA Extraction System (EuroClone S.p.A., Olaszország) kett alkalmazásával.

A tisztított DNS-t Illumina HiSeq2000 platformon szekvenáltattuk, amelynek eredményeként 4 paired-end és 2 mate pair szekvencia könyvtár készült el. Ezekben a könyvtárakban összesen több mint 2 milliárd nyers read szekvencia

található, ami 223 milliárd bázispár hosszú DNS-nek felel meg. Ez a gímszarvas haploid genomra vonatkoztatva 74-szeres átlagos lefedettséget jelent.

Az egymást átfedő paired-end readekből az ALLPATHS-LG assembler program először összefüggő, kontinuos szekvenciákat contigokat hozott létre. A contigokból pedig még nagyobb szekvencia egységeket scaffoldokat generált. Azokon a részeken, ahol nincsenek átfedő readek a mate pair read párok tagjai adták meg az azonos scaffoldra eső contigokat. Összesen 437412 darab contig keletkezett, amelyek N50 értéke 7.5 Kbp. A contigokból összefűzött scaffoldok száma 34724, N50 értékük 430 Kbp volt. A munkám során a kettős referencia vezérelt genom összeállítás módszerét alkalmaztam. Az első referenciát a gímszarvas kapcsoltsági géntérképe jelentette, amelyet két új-zélandi szarvas farmon interspecifikus back-cross keresztezési eljárással állítottak elő. 7 darab F1 milu (*Elaphurus davidianus*) és gímszarvas (*Cervus elaphus*) hibrid szarvasbikát kereszteztek 267 gímszarvas ünővel, amelyek eredményeképpen 351 back-cross utód született. A back-cross szarvasokra készített géntérképen 714 markert határoztak meg, amelyekből azonban csak 621-et tudtak pontos térképpontként elhelyezni. A markerek típusaik alapján lehettek mikroszatelliták, RFLV-k, EST-ek, proteinek, és AFLP-k. A géntérkép a gímszarvas haploid kromoszóma számának megfelelően az X/Y kromoszómákkal együtt 34 kapcsoltsági csoportba rendeződik és 2532 cM hosszúságú.

A másik referenciát egy közel rokonfaj a szarvasmarha (*Bos taurus*) online elérhető fizikai géntérképe, jól annotált referencia genom szekvenciája (Btau_5.0.1) adta.

2. CÉLKITŰZÉSEK

A fő célom az volt, hogy a Gímszarvas/Csodaszarvas Genom Program keretén belül létrehozzam a „Gímszarvas Referencia Genom szekvenciát (CerEla1.0), ami a világ első gímszarvas, és egyben a Magyarországon készült első, teljes, emlős referencia genom szekvenciája. A munkám során a következő részcélokat tűztem ki.

1. Megfeleltetni egymással a rendelkezésre álló gímszarvas genetikai térképet (Slate és mtsai., A Deer (Subfamily Cervinae) Genetic Linkage Map and the Evolution of Ruminant Genomes, 2002) és a *de novo* gímszarvas szekvencia vázakat (genom assembly/scaffoldokat).
2. Egyeztetni egymással a gímszarvas genetikai térkép pontjait (DNS markereit) és a szarvasmarha referencia genom ortológ szekvenciáit.
3. Azonosítani a szarvasmarha referencia gének ortológ szekvenciáit a gímszarvas scaffoldokon.
4. Összehasonlítani és felilleszteni a *de novo* gímszarvas scaffoldokat a szarvasmarha referencia genomra az evolúciós változásokat figyelembe véve.
5. Megkeresni a gímszarvas genom fehérje kódoló génjeit, repetitív szekvenciáit, transzfer RNS, riboszómális RNS és mikro RNS szekvenciáit.
6. Meghatározni a gímszarvas kromoszómák centromeron pozícióit.
7. Az egyedi azonosítóval ellátott gímszarvas teljes genom feltöltése és online elérhetővé, letölthetővé tétele az NCBI szerverén.

3. ANYAG ÉS MÓDSZER

A munkám első lépéseként a gímszarvas géntérkép pontjainak DNS szekvenciáit név és azonosító alapján irodalmi adatokból és online elérhető Ensembl, UCSC, Uniprot, NCBI, és ENA adatbázisokból gyűjtöttem ki. Az összes gímszarvas scaffoldból és a szarvasmarha referencia genomából makeblastdb paranccsal külön-külön blast könyvtárakat készítettem. A térképponti markereket és a 10 darab DeerPlex mikroszatellitát úgynevezett query vagyis kereső multifasta fájlomba fűztem. A markerek (query) illesztése a szarvasmarha és a gímszarvas-scaffold blast könyvtárakra blastn, tblastx, és tblastn paranccsal történt. Az ilyen módon kihalászott scaffoldokat „mapmarker vagyis térképpont” scaffoldoknak (MMSc) neveztem el. A továbbiakban a gímszarvas térkép pontok megfelelőit is azonosítottam a szarvasmarha teljes genom szekvenciában (összehasonlító géntérképezési elv). A gímszarvas géntérkép pontjai azonos sorrendben, kolineárisan helyezkedtek el a szarvasmarha genomban is, azaz kiterjedt szinténiák, lokális kapcsoltságok figyelhetők meg.

A két faj marker szekvencia hasonlósága és szinténiái miatt feltételezhető a génjeik közötti hasonlóság is, ezért a következő lépésben MegaBLAST paranccsal a UCSC online adatbázisból letöltött szarvasmarha géneket illesztettem a gímszarvas scaffold blast könyvtárhoz. Az ilyen módon kihalászott scaffoldokat reference gene containing scaffolds-nak (RGSc) neveztem el.

A referenciaként használt szarvasmarha genomról RepeatMasker programmal kimaszkoltam (hard-masking) a kis komplexitású genomi szekvenciákat és a repetitív elemeket, hogy ne zavarják a későbbi illesztéseket. Ezután a referenciát seqretsplit paranccsal kromoszómákra szedtem szét. A gímszarvas és a szarvasmarha kromoszómák száma eltér, és az evolúciójuk során kromoszóma átrendeződések (Robertsoniális fúzió, Robertsoniális transzlokáció, hasadás, transzlokáció, paracentrikus inverzió, és inverzió) történtek. A két faj kromoszómális különbsége miatt a referenciaként használt szarvasmarha kromoszómákat a gímszarvas géntérképnek megfelelően alakítottam át a seqret

program segítségével. Néhány esetben két szomszédos marker megcserélődését figyeltem meg, ilyenkor az 1 cM-nál, vagy annál nagyobb távolság esetén elképzelhetőnek tartottam a valós inverziót, azonban 1 cM alatt megmaradt a szarvasmarhának megfelelő marker rangsor.

A markerek és az MMSc-k szarvasmarha genomra illesztéséhez a BLAST, a MUMmer3.0, a BWA 0.7.10-r789 verzióját futtattam le.

Az RGSc-k szarvasmarha kromoszómális pozícióinak meghatározása szintén MegaBLAST, MUMmer és BWA programokkal történt.

Ezután a nem referencia gének (a USCS Refseq génjein kívüli géneket (miRNS, tRNS, rRNS gének) tartalmazó scaffoldok vagyis az úgynevezett inter reference genes scaffolds-ok, IRGSc-k) felillesztéséhez a MegaBLAST, a LASTZ_32, a MUMmer, és a BWA bioinformatikai szoftver csomagokat használtam. Ennél a munkafolyamatnál az RGSc-k és az IRGSc-k kromoszómális lokalizációját, sorrendjét és irányultságát határoztam meg a gímszarvas kapcsoltsági csoportoknak megfelelően széttördelt, és hard repeatmaszkolt szarvasmarha kromoszómákon. A többféle program eredményfájljait házilag készített Bash szkriptekkel dolgoztam fel, lényegében awk, sed, bedtools merge és sort parancsokkal.

A szarvasított (gímszarvas kapcsoltsági csoportoknak megfelelően széttördelt) szarvasmarha referencia genomnak az MMSc, RGSc és IRGSc scaffoldokkal lefedett részeit kimaszkoltam (hard masking). A megmaradt réseket a már ismertett BWA és MUMmer programok alkalmazásával feltöltöttem a kimaradt 1999 bp feletti scaffoldokkal (GFSc, gap filling scaffolds).

A megfelelő sorrendben elhelyezett scaffoldok, és 2582 egyedi contig összefűzése Python szkripttel történt úgy, hogy a közöttük lévő résekbe 100 bp-nyi N karakter került.

A repetitív szekvenciákat a RepeatMasker Open-4.0. nevű szoftverrel kerestem meg. Az emlős riboszómális RNS kis és nagy alegység (SSU, LSU) szekvenciáit a SILVA123 referencia adatbázisból szedtem ki. A prekursor formátumú

mikroRNS-ek a miRBase adatbázisból származtak, amelyekből 21 emlős miRNS szekvenciát használtam fel a későbbiekben. Az rRNS-ekből és külön a miRNS-ekből kereső-szekvencia fájlokat hoztam létre, amelyeket blastn paranccsal illesztettem a gímszarvas genomunkhoz. A riboszómális RNS 5S egységét Barrnap 0.6 program segítségével kerestem meg. A transzfer RNS-eket a tRNAscan-SE-1.3.1 szoftverrel találtam meg.

A fehérje kódoló géneket a MAKER 2.31.8 gén annotációs program azonosította, olyan módon, hogy egymás után hívott be különböző alprogramokat (RepeatMasker-open-4.0.5, RepeatRunner, RepeatModeler 1.0.4, RECON, RepeatScout 1.0.5, Exonerate, BLAST, SNAP, AUGUSTUS). A MAKER más fajokból származó EST, mRNS és fehérje szekvenciákat (szikaszarvas, szarvasmarha, juh, ember) illesztett fel BLAST algoritmussal a könnyű maszkolt (soft-masking) gímszarvas genomra. Az *ab-initio* gén predikciót pedig a SNAP és az AUGUSTUS programokkal végezte el. A fehérje funkciók és a protein-kódoló gének azonosítása az InterProScan szoftverrel történt.

A gímszarvas genetikai variánsok (SNV, INDEL) meghatározásánál a paired-end és a mate pair könyvtárak readjeit BWA -mem program illesztette fel referencia genomra. A felillesztett readekben a SAMtools (paraméterek: mpileup -D -S -E -uf) szoftver azonosította a genetikai variánsokat. Az aminosavak funkcionális változásait okozó variánsok annotációját ANNOVAR szoftver végezte el.

4. EREDMÉNYEK

A gímszarvas genom program első lépéseként egy hétéves, Bószénfán nevelkedett, kapitális, később 12 kg súlyú agancsot viselő gímszarvasbika véréből DNS-t izoláltunk. Az állat teljes genomját szekvenáltattuk Illumina HiSeq2000 szekvenáló platformon. E folyamat következményeként 4 paired-end és 2 mate pair szekvencia könyvtár készült el, amelyekben összesen körülbelül $2,2 \times 10^9$ read található, amely 222,7 Gbp DNS-nek felel meg. A *de novo* assembly készítést az ALLPATHS-LG program hajtotta végre, melynek eredményeképpen 437412, összesen 1,95 Gbp hosszú contig jött létre. A contigokat a program nagyobb szekvencia egységekké, scaffoldokká fűzte össze. A scaffoldok száma 34724, össz hosszúságuk pedig a gapeket is figyelembe véve 3,4 Gbp.

Az első számú referenciaként használt gímszarvas genetikai térképe (Slate és mtsai., A Deer (Subfamily Cervinae) Genetic Linkage Map and the Evolution of Ruminant Genomes, 2002) az összes kapcsoltsági csoportra nézve 2532 cM hosszúságú, 34 kapcsoltsági csoportból áll és 5,7 cM-os marker átlagsűrűséggel rendelkezik. A linkage groupokon meghatározott 621 genetikai markert online genomikai adatbázisokból kerestem ki. Az AFLP-ék kivételével az összes típushoz (EST, RFLV, STS, protein) találtam szekvenciákat. Ekképpen 365 marker DNS szekvenciáját sikeresen azonosítottam és helyeztem el a genetikai térképen. A 365 markert hordozó scaffoldokat „mapmarker” (MMSc) scaffoldoknak neveztem el. Az MMSc-k így meghatározott pozícióba kerültek a kapcsoltsági géntérképen.

Második számú referenciaként a szarvasmarha referencia genom (NCBI Btau_5.0.1) szolgált. Erre a genomra illesztettem fel a 365 *C. elaphus* marker szekvenciáit, illetve térképpont scaffoldjait (MMSc). Így sikeresen azonosítottam a *B. taurus* genomban a 365 gímszarvas MMSc-vel hasonlóságot mutató, ortológ szekvenciákat. Ezek pozíciói a kromoszómákon végig kolineárisan helyezkedtek el mindkét fajban. Ezenfelül az MMSc-ben található gének sorrendje is azonos volt a „bovin” génekével, vagyis “intra-scaffold és intra-contig” szinten is érvényesült a szinténia. A gímszarvas és a szarvasmarha MMSc szekvenciák kolinearitása

okán feltételezhető, hogy a két faj génszekvenciái is nagyon hasonlóak. A feltételezett hasonlóság miatt szarvasmarha géneket töltöttem le online elérhető UCSC genomikai adatbázisból.

A továbbiakban ezekkel a kereső génekkel, mint „csalikkal” kihalásztam azokat a „préda” scaffoldokat, amelyek tartalmazták a csali génnel ortológ gímszarvas szekvenciákat. Az ilyen scaffoldokat „referencia gén scaffoldoknak” (RGSc-knek) neveztem el. Annál bizonyosabb volt egy RGSc helyzete minél több, azonos sorrendben következő szarvasmarha gén ortológ szekvenciát foglalt magában. A folyamat eredményeként a markerpontok közötti részekre RGSc-eket helyeztem el. A gímszarvas genetikai térkép szegmensekre igazított bovin génsorrendjét követve összesen 6013 scaffoldot lokalizáltam.

A következő munkafázisban az MMSc-RGSc és az RGSc-RGSc közötti réseket töltöttem fel az eddig el nem helyezett scaffoldokkal. Ezek feltehetően nem tartalmazták az UCSC által megadott Refseq proteinkódoló géneket, de nagy valószínűséggel rendelkeztek másmilyen fehérje kódoló, például nem Refseq, rRNS, tRNS, és miRNS génekkel. Ezeket a kereső gímszarvas scaffoldokat, „inter referencia gén scaffold”-oknak (IRGSc) neveztem el. Az MMSc-eket, RGSc-eket és IRGSc-eket együttvéve 13748 scaffoldot tudtam elhelyezni, azonban 15205 2 Kbp-nál hosszabb scaffold lokalizációja továbbra is kérdéses maradt.

A probléma megoldásaként kimaszkoltam a már meglévő 13748 *Bos taurus* genomi helyet. Az ilyen módon „maszkolt” referenciára illeszttem fel a megmaradt 2 Kbp-nál nagyobb szekvenciákat. Ennélfogva még 9845 új, úgynevezett réskitöltő, vagyis „gap filling” scaffoldhoz (GFSc) találtam szarvasmarha kromoszóma részeket. A munkafolyamat során sikerült azonosítani összesen 23593 scaffold szarvasmarha ortológ régióját.

A szekvenciák nagyrésznél, pontosabban 99,6%-ánál egyértelmű, kölcsönös megfeleltetést tapasztaltam a két faj között, de 102 esetben elmondható volt, hogy akár több, egymástól elkülönülő kromoszómális szakasszal mutattak ortológiát vagy ezzel ellenkezőleg, azonos pozíciókat fedtek le. Megoldandó a problémát

önálló contigokra bontottam a scaffoldokat vagy összefésültem őket egy azonos scaffoldban. A keletkezett 2582 egyedi contigot és az új 35 scaffoldot ebben a formában illesztettem újra a referencia genomhoz.

A pszeudokromoszóma készítés végeredményeként létrehoztam a haploid gímszarvas (*C. elaphus hippelaphus*) referencia genomot, amely CerEla1.0 nevet kapta. A CerEla1.0 genom 23491 plusz 35 scaffoldból (MMSc-k, RGSc-k, IRGSc-k, GFSc-k plusz a 35 új) és 2582 önálló contigból tevődik össze, azaz összegezve 26108 szekvencia elemet tartalmaz, ami 3,4 Gbp genom hosszúságot ad ki.

Az „unplaced” kategóriába került 11444 scaffold, mivel nem tudtam az új gímszarvas referencia kromoszómákon elhelyezni őket. Az unplaced szekvenciák összesen 52989442 bp-t tesznek ki, ami a teljes genom 1,6 %-a.

Nagy jelentőséggel bírtak azok a scaffoldok, amelyek több gént, vagyis lehetséges kapcsolt, szinténikus genetikai elemet tartalmaztak. A Btau_5.0.1 szarvasmarha genommal a gímszarvas kromoszómákat összevetve kiderült, hogy az intra-scaffoldikus gének a szarvasmarha ortológ génekkel megegyezően lokális kapcsoltságokat mutattak. Minden több génes gímszarvas scaffold (3422) és ortológ bovin kromoszómális szegment esetében észleltem ezeket a kiterjedt szinténiákat. Elmondható tehát, hogy a szinténia és a gének sorrendje nem csupán térképpont markerekre kromoszómális szinten igazolható, hanem szubkromoszómálisan, a scaffoldokon belül is megnyilvánul.

A gímszarvas gének identifikációja a szarvasmarha, juh és emberi transzkriptom és proteom illetve a *de novo* létrehozott szikaszarvas (*Cervus nippon*) transzkriptom felhasználásával, a MAKER program alkalmazásával történt. A „pipeline”-okat futtató MAKER 19368 fehérje kódoló gént azonosított. Az identikus gének sorrendje túlnyomórészt megegyezett a szarvasmarha és a gímszarvas genomban. A teljes referencia genom (CerEla1.0) felkerült az NCBI-ba (NCBI azonosító MKHE00000000.1, GCA_002197005.1), ahol genom böngészővel megjeleníthetők az annotált gének.

Ezenfelül sikerült meghatározni 589 rRNS kódoló gén (LSU, SSU) helyzetét,

amely 98.3 Kbp hosszúságú szekvenciát fedett le, ami a pszeudokromoszóma összhossz 0,0029%-át tette ki. Az 5s rRNS génekből 1029-et (96Kbp, a CerEla1.0 0,0028%-a), a tRNS-ekéből 2096-ot (128 Kbp, a CerEla1.0 0,0038%-a) a mikroRNS-ekéből pedig 264-et (27,7 Kbp, a CerEla1.0 0,0008%-a) lokalizáltam. A RepeatMasker program 769492957 bp hosszúságú repetitív szakaszt ismert fel, ami a teljes genom 22,73% -át adja ki.

A *C. elaphus* kromoszómák felépítése meglehetősen „primitív” jelleget mutat, mivel szinte az összes akrocentrikus (A kromoszóma). A gímszarvas kromoszómák többsége (19 autoszóma és X, Y) közvetlenül párba állítható egy homeológ szarvasmarha kromoszómával. A szarvasmarhában az evolúció során megőrződött a 6 akrocentrikus ősi kromoszóma, és ezek hasadásával a szarvasfélék evolúciója alatt keletkezett 12 akrocentrikus (szarvas) kromoszóma. Ez azt jelenti, hogy a szarvas evolúciós vonalon új centromeronok jöttek létre. A Ce5 metacentrikus (M kromoszóma) a *Cervinae* ág leválásakor két akrocentrikus kromoszóma Robertsoniális transzlokációjával jött létre. A *Pecora* leszármazási vonalak (*Bovidae*, *Cervidae*) kariogramjai alapján, az evolúciójuk során nagyon gyakran történt tandem összekapcsolódás úgynevezett Robertsoniális fúzió és hasadás, azaz amikor 2 akrocentrikusból 1 akrocentrikus kromoszóma képződik és vice versa. A gímszarvas genetikai térkép nem jelezte a centromeronok pozícióját. A kapcsoltsági csoportok pontjait itt nem lehet egységesen a centromeronok helyzetéhez orientálni. A szarvasmarha oldalán a genom szekvenciákban jelezték a centromeronok helyzetét, viszont korábbi kromoszóma citológiai vizsgálatokban centromeron festéssel és in situ DNS hibridizációk alapján meghatározták a centromeronok és a hozzájuk közel eső gének lokalizációját. Ezek a gének már azonosíthatók voltak a szarvasmarha genom szekvenciában is, ortológjaik pedig szintén szerepeltek a gímszarvas genom szekvenciában, CerEla1.0-ban. A szarvasmarha genomban azonosított „centromeron-közeli gén” szinténiákat átvezettem a homeológ gímszarvas kromoszómákra. A rendelkezésre álló, egymással ortológ szarvasmarha és gímszarvas pszeudokromoszóma szekvenciák

és citogenetikai felvételek (a sorba rendezett metafázisos kromoszómák sávmintázatai) komparatív elemzésével sikerült tehát meghatározni a centromeronok lehetséges helyzetét (a géntérképi pontokhoz igazítva) mind a 34 gímszarvas kapcsoltsági csoporton és a genetikai térképek alapján készített pszeudokromoszómákon. Ilyen módon 6 szarvasmarha kromoszómával (Bt5, Bt6, Bt2, Bt8, Bt1, Bt9) párosan megfeleltethető 12 gímszarvas kromoszómán (Ce3-22, Ce6-17, Ce8-33, Ce16-29, Ce19-31, Ce26-28) is nagy biztonsággal meg lehetett adni a centromeronok helyzetét. Az akrocentrikus gímszarvas Ce19, a szarvasmarha Bt1 disztális felével, a szintén akrocentrikus gímszarvas Ce31 viszont a szarvasmarha Bt1 proximális szakaszával ortológ. A helyzet azonban összetettebb ennél, hiszen a Ce19 kromoszóma ősében egy jól definiálható törés és transzlokáció (az alsó és a felső szegmens helyet cserélt) is lejátszódott az evolúció során. A gímszarvas Ce28 és a Ce26 akrocentrikus kromoszóma a szarvasmarha Bt9 akrocentrikus kromoszóma két karjával azonos. Ennek az az oka, hogy a szarvasfélék ősében ez a kromoszóma kettéhasadt. A gímszarvas Ce26 centromerájának nincs megfelelője a szarvasmarha Bt9 szekvenciáiban, mivel gímszarvas Ce28 „örökölte meg” a centromeront, de a sávmintázatok és a genetikai térképpont sorrend összevetése alapján a gímszarvasban le kellett játszódnia egy paracentrikus inverzióknak a fajképződés folyamán. Ennek következtében a kromoszóma összes ortológ markereit érintő, nagyméretű szegmense egészben megfordult, míg a centromeronhoz legközelebbi szekvenciák a helyükön maradtak. Két olyan esetet tártam fel, amikor egy gímszarvas kromoszóma két szarvasmarha kromoszómával ortológ. A metacentrikus gímszarvas Ce5 egyik karja a szarvasmarha Bt17-el a másik a Bt19-el egyezik meg. Az akrocentrikus gímszarvas Ce15 Robertsoniális fúzióval keletkezett, amelyben a centromeron közeli rész a szarvasmarha Bt28-al, a távoli rész pedig a Bt26-al ekvivalens.

Az inverziók, a transzlokációk, a kromoszómális fúziók és törések áthelyezik és elkülönítik a térképponti markereket és új környezetet hoznak létre az átrendeződések töréspontjai körül. A gímszarvas-szarvasmarha rokoni kapcsolatát

tekintve ezek az "evolúció által létrehozott" átrendezések új szomszédos szekvencia régiók kialakulásához vezettek. A CerEla1.0 genomban 26 darab szarvas-bovin átrendezést észleltem, amely 18 inverziót, 2 transzlokációt, és 6 kromoszómális fúziót és hasadást foglalt magába. Abban a 6 esetben, amikor 1 akrocentrikus szarvasmarha kromoszóma két akrocentrikus gímszarvas kromoszómával feleltethető meg az ortológ szarvasmarha régióhoz illeszkedő scaffoldokat és contigokat úgy osztottam fel a két gímszarvas kromoszóma között, hogy figyelembe vettem a gímszarvas genetikai térképeken található rekombinációs távolságokat és ezeknek megfelelően, arányosan, illetve a szarvasmarhában tapasztalt sorrend szerint rendeztem el őket. Az ilyen típusú kombinált DNS szekvenciák a CerEla1.0 genom 5 %-át (0,166 Gbp) teszik ki. Az inverziók 54 gímszarvas-szarvasmarha szekvencia váltás („switch pont”) pontot eredményeztek, amelyek a túlnyúló, „flanking” régióikkal együttvéve tulajdonképpen megegyeztek a szomszédos MMSc-k által határolt, összesen 462,9 Mbp szekvenciával. Ez a CerEla1.0 genom 13,5 %-át jelenti. A CerEla1.0 81,5% -ában a gímszarvas gének az MMSc közötti szegmensekben a szarvasmarha ortológ gének sorrendjét követik, az MMSc-ken belül pedig egyértelműen a gímszarvas génsorrend érvényesült. A CerEla1.0 szekvencia 18,5%-ban (kromoszómális hasadások/fúziók, flanking részek inverziói) gímszarvas és szarvasmarha génjeinek szinténikus blokkjai kombinálódtak.

A CerEla1.0 genom heterozigocitás/genetikai variáns keresése során 2807458 SNV-t és 364689 indelt azonosítottunk. A MAKER annotációs pipeline eredmény fájljai alapján további a heterozigóta SNV-eket annotáltunk. Ily módon összesen 17700 nem-szinonim és 14252 szinonim SNV-t találtunk meg.

* Bt(valamilyen szám): *Bos taurus*, vagyis szarvasmarha valamilyen számú kromoszómája, Ce(valamilyen szám): *Cervus elaphus*, vagyis gímszarvas valamilyen számú kromoszómája

5. KÖVETKEZTETÉSEK ÉS JAVASLATOK

Az értekezés témája kiemelkedően fontos a hazai genomikai kutatás szempontjából, hiszen Magyarország első, valódi, nemzetközileg is elismert teljes referencia genom szekvenciája előállításáról szól. Nagymértékben hozzájárul a gímszarvas populációgenetikai profiljának megismeréséhez, ezáltal lehetőséget teremt új mikroszatellita és SNP markerek tervezésére. A gímszarvas referencia genom segítheti e nagyvad evolúcióbiológiájának, leszármazási vonalainak a megismerését, és a farmtenyésztési eljárások fejlesztését. A gímszarvas gének szerkezetének és azok promóter szekvenciáinak ismerete alapján feltárhatjuk a kapitális agancossal rendelkező egyedek genetikai hátterét. Feltérképezhetőek lesznek a csont-agancs metabolizmusban és tumor képződésben résztvevő gének működései és orvosbiológiai vonatkozásai. A disszertációban ismertetett bioinformatikai módszerek megfelelő mintául szolgálhatnak más mezőgazdasági vagy környezetvédelmi szempontból értékes faj referencia genomjának a létrehozásához.

A bioinformatikai munkát megelőzte a mintagyűjtés, a DNS izolálás és a teljes genomi DNS szekvenálása. A DNS mintavétel a Kaposvári Egyetem Vadgazdálkodási Tájéközpont Bőszenfai Szarvas farmján élő 7 éves, kapitális gímszarvasbika véréből történt az állatvédelmi törvények figyelembevételével. A DNS izolálás eredményeképpen keletkezett nagy tisztaságú DNS-t a dániai Aros Applied Biotechnology cég szekvenálta meg Illumina HiSeq2000 készülékkel, aminek következményeként nagy mennyiségű, további vizsgálatokra alkalmas, bioinformatikai nyers adat 2 milliárd read szekvencia jött létre. A későbbi populációgenetikai vizsgálatok szempontjából érdemes lenne több, különböző magyarországi populációból származó körülbelül 150-300 gímszarvas egyedre kiterjeszteni a mintavételt és a teljes genom szekvenálást. A szekvenálási módszer tekintetében ajánlatos az Illumina technológiánál maradni, hiszen új technológiákkal a leolvasott read méretének növelésére és az eljárás költségének a csökkentésére törekszenek.

A Broad Institute ALLPATHS-LG programja a read szekvenciákat contigokká és scaffoldokká fűzte össze. A scaffoldok száma 34724 volt. Viszonylag sok 2000 bp alatti, kisméretű és kevés nagy méretű scaffold jött létre. Amelyekből a kromoszómába rendezés szempontjából a 8 Kbp fölöttiek bizonyultak a legjobban használhatónak. A scaffoldok N50 érték-e 430 Kbp a teljes hosszukra nézve, és N50=265 Kbp értéket kaptunk a gap-mentes scaffoldok esetében.

A referenciaként használt *C. elaphus* géntérkép hosszúsága 2532 cM, míg az „összeszerelt” CerEla1.0 referencia genom összes hosszúsága 3,4 Gbp, vagyis 1 cM 1,34 Mbp-nak felel meg. Ez az érték szignifikánsan nagyobb, mint az általánosan használt megközelítő 1cM/1Mbp érték (a „hüvelykujj mérték/zsinórmérték, angolul thumb rule”) vagy a szarvasmarha genomra megállapított 0,8 Mbp/1cM. A gímszarvas CerEla1.0 referencia genom látszólagosan 25%-kal hosszabbnak adódik, mint a szarvasmarha, Btau_5.0.1 referencia genom. A gímszarvas genom 0,7 Gbp-nyi „többlet” hosszúságának az okát keresve elvégeztem a CerEla1.0 és Btau_5.0.1 pszedogenom ortológ szegmenseinek összehasonlítását. Azt tapasztaltam, hogy néhány kivételtől eltekintve a gímszarvas és a szarvasmarha pszedogenomok mentén a gímszarvas szegmensek és végsősoron a scaffoldok egyenletesen 1,25-ször hosszabbak, mint a szarvasmarha szegmensek (Kivétel ez alól a Ce11 kromoszóma, ahol ez az arány 2,2). Összevetve a 25 %-os (0,7 Gbp) extrahosszúsággal rendelkező CerEla1.0 pszedogenomot (3,4 Gbp) a Btau_5.0.1 pszedogenommal (2,7 Gbp) megállapítható, hogy túl sok gap régiót („NNN”) épített be az ALLPATHS-LG program a contigok közé a scaffoldokban. Feltéve, hogy a *B. taurus* és *C. elaphus* genomjai méretüket tekintve lényegében megegyeznek egymással, továbbá figyelembe véve, hogy a CerEla1.0-ban található a contigok össz DNS-szekvenciái 1,9 Gbp-t tesznek ki feltételezhető, hogy a szarvasmarha genommal arányosítható 0,8 Gbp gap régió helyett 1,5 Gbp-nyi gap került a scaffoldokba összesen. Ezek eloszlása azonban a fizikai távolságoknak megfelelően arányos. A szarvas szarvasmarha szinténikus szegmens hosszaknál tapasztalt 1,25 arány a Ce11

kromoszóma esetén lényegesen több, 2,2 volt ennek magyarázatára nincs javaslat. A gímszarvas scaffoldokban a térképponti markerek szegmensei és gének sorrendje szinténiát mutatott a közel rokon faj szarvasmarha ortológ szegmenseivel, vagyis a scaffoldok lényeges, nukleotid szekvenciái mégis jónak bizonyultak. További gímszarvas egyedek assembly készítésekor megfontolandó más ALLPATHS-LG paraméterek vagy újabb programok (például DISCOVAR) használata.

A scaffoldok kromoszómába rendezése Jon Slate gímszarvas genetikai marker térkép és a szarvasmarha referencia genom felhasználásával történt meg. A gímszarvas genetikai térképén 621 EST, RFLV, STS, protein genetikai markert határoztak meg, ám ezek szekvenciáját irodalmi adatokból és bioinformatikai adatbázisokból kellett kigyűjteni. A problémát az jelentette, hogy 229 AFLP-hez sehol sem lehetett szekvenciát találni, tehát ezek kiestek a további vizsgálatokból, és a többi kategóriánál is előfordult, hogy ismeretlen maradt egy-egy marker szekvenciája. Ezen okok miatt 365 markerpont DNS szekvenciáját azonosítottam és a továbbiakban kereső-szekvenciaként BLAST programmal illeszttem a scaffoldokhoz és a szarvasmarha referencia genomhoz. Az ilyen módon megtalált scaffoldokat térképpont vagyis "mapmarker" scaffoldoknak (MMSc-k) neveztem el. A vizsgálathoz több referencia genom verziót (NCBI Btau_4.6.1, Btau_5.0.1) is használtam, bár az akkoriban legfrissebb is elegendő lett volna (Btau_5.0.1). A gímszarvas géntérkép pontjai hosszú szakaszokon azonos sorrendben, kolineárisan helyezkedtek el a szarvasmarha genomban és a gímszarvas scaffoldokban egyaránt, emiatt az összehasonlító géntérképezési elvet felhasználva a szarvasmarha ortológ génekkel kihalászott gímszarvas scaffoldokkal (RGSc) töltöttem fel a marker pontok térképközeit. A későbbiekben ellenőrzésképpen a géneket tartalmazó scaffoldokat és a többi, 2 Kbp-nál hosszabb, referencia gént nem tartalmazó 15205 scaffoldot illeszttem fel a szarvasmarha referencia genomra LASTZ, MUMmer, BWA és MegaBLAST programok alkalmazásával. Ezeket neveztem el inter-referencia gén scaffoldoknak

(IRGSc).

A gímszarvas genomon lefutott MAKER program annotációs folyamata során 19368 protein-kódoló gént talált, amely teljesen megfelel a többi emlős fajban is leírt 19000 és 21000 közé eső fehérje kódoló gén számnak és a kérődzők ismert génjeinek 90 %-át lefedi. A gímszarvas gének és szarvasmarha ortológjaik sorrendje kolinearitást mutat a CerEla1.0 és a Btau_5.0.1 genomban. Meghatároztuk a gének részletes szerkezetét (exon, intron szerkezete, 5' és 3' UTR régióik) evolúciós rokonságait (hasonló gének más fajokban), funkcióit (a sejten, szerveken, szöveteken belüli, fejlődésben betöltött szerepüket). Feltártuk az ismétlődő, repetitív szekvenciákat, LTR elemeket, transzfer RNS, kis RNS és riboszóma RNS géneket, és mindezen genetikai elemek helyzetét a kromoszómákon. Azonosítottunk 2,8 millió heterozigóta pont változatot (SNV) és 365 ezer apró deléciót és inszerciót a gímszarvas genomban.

A CerEla1.0 gímszarvas genomtól teljesen függetlenül, egy angol-újzélandi kutatócsoport által meghatározott, több mint 38 ezer nagyrészt gímszarvas kisebb részt szikaszarvas SNP rekombinációs pont elemzését tartalmazó nagy sűrűségű gímszarvas genetikai térkép is elkészült. Ez a géntérkép nagyszámú, 38083 SNP marker szekvenciát tartalmaz, amely egy komparatív vizsgálattal validálja a magyar gímszarvas referencia genomot, a CerEla1.0-t. Az összehasonlító munka első lépéseként kikerestem és letöltöttem az összes angol-újzélandi SNP markert a 150-150 bázispáros szárnyi (flanking) régióikkal. Ezeket a markereket, mint kereső szekvenciákat a mi gímszarvas genomunkra illesztettem fel BLASTN paranccsal. Kiválasztottam a legjobb találatokat, amelyek eredményeiből táblázatokat generáltam minden egyes autoszómára olyan módon, hogy látható legyen a mi gímszarvas kromoszómális pozíciónk megabázispárban megadva és a kereső angol-újzélandi SNP marker megabázis és cM pozíciója. A kapott eredményekből pont diagramokat szerkesztettem excel programmal. Az adatok és ábrák alapján jól láthatók a lehetséges töréspontok és kromoszóma átrendeződések. A 38083 pontból 32408-at tudtam elhelyezni az elvártnak

megfelelő kromoszómán, vagyis az összes pont 85%-át. Ezen pontoknál 33 esetben tapasztaltam több pontból álló rész nagyobb inverzióját. A fordulások legvalószínűbb oka a Slate-féle 621 marker és a Johnston-féle SNP marker genetikai távolság alapú géntérképek között fennálló különbség, illetve, hogy két fordított helyzetű, szomszédos Slate marker esetében nem lehettem biztos abban mekkora flanking részek tartoznak az invertált szakaszokhoz. Továbbá a gímszarvas referencia genom elkészítésénél a Slate térkép egyes szomszédos markerpontjai fordított helyzetűnek tűntek a szarvasmarha referencia genomhoz képest is, ami 54 gímszarvas-szarvasmarha inverziót, összesen 462,9 Mbp szekvenciát adott. A CerEla1.0 létrehozásához a Slate-féle géntérkép állt a rendelkezésre, tehát a scaffoldok elhelyezése e sorrendiségnek megfelelően történt. A felillesztett pontokból csupán 604 adott kiugró értéket, ami kevesebb, mint az elhelyezett pontok 1,9 %-a. 1019 SNP -t más kromoszómán lokalizált a BLAST program. Ezen pontok diszkrepanciáját az okozhatta, hogy a gímszarvas pszeudokromoszóma készítéshez használt közel rokonfaj, a szarvasmarha kromoszómaszáma nem egyezik meg a gímszarvaséval. Azokban az esetekben amikor 12 gímszarvas 6 szarvasmarha kromoszómával ortológ a genom összeállítás során nem lehetett biztosan tudni, hogy az utolsó Slate-féle genetikai térképpontokon túl a két gímszarvas kromoszóma miként „osztózik” az ortológ szarvasmarha kromoszóma „elosztandó” szegmensén. Ezek a bizonytalan részek 0,166 Gbp hosszú DNS szekvenciát tesznek ki. Összességében azonban elmondható, hogy a kapott eredmények meggyőző bizonyítékot szolgáltatnak a CerEla1.0 gímszarvas genom helyességének igazolásához.

A nagysűrűségű SNP genetikai térképen kívül a munka kezdetétől fogva rendelkezésre állt a rendkívüli érzékenységű, 10 autoszomális tetranukleotid mikroszatellita lokuszra épülő „parentage controll kit”, a DeerPlex. A vizsgálat során sikerült a DeerPlex mikroszatellitákat kiterjedt DNS szekvencia környezetben elhelyezni.

A CerEla1.0 felhasználható SNP és/vagy mikroszatellita alapú egyedazonosításhoz, populációgenetikai szintű vizsgálatokhoz, leszármazási vonalak és evolúciós viszonyok feltáráshoz. Jelenleg publikálás alatt van egy cikkünk, amelyben új XY kromoszómás mikroszatellita markerek fejlesztéséről írunk. Ezek a nemi kromoszómás markerek nagyban megkönnyítik az apai és az anyai leszármazási vonalak feltárást, lehetővé teszik az objektív és egyértelmű egyedazonosítást, a tudatosabb és megbízhatóbb állatnemesítést és tenyésztést, valamint megbízható eszközt jelenthetnek a nagymértékű vadorzás visszaszorításában. További orvosbiológiai témájú cikkek alapjául szolgálhatnak a gímszarvas csont és agancs metabolizmus gének promóter szekvenciáinak az elemzése. Az éves agancs ciklusuk során a gímszarvas bikák februártól májusig lehullatják csontos agancsukat, ezután 100-120 napig tart az új agancs épülése. Ilyenkor nagymennyiségű ásványi anyagot kell biztosítani és eljuttatni az akár 14-17 kg-os csontszervbe. A tápanyag kalcium bevitel kevés, ezért vázelemekből-szegycsontból, bordákból, egyes csigolyákból transzportálják a kalciumot és foszfátot az agancs mineralizációjához, ezzel csontsűrűség csökkenést, azaz fiziológias csontitkulást indukálva az érintett vázcsontokban. Később az ásványi sók a dús vegetációból táplálkozás útján visszapótlódnak, azaz a csontsűrűség helyreáll a csontvázban. Kutatócsoportunk előzetes vizsgálatai szerint 8 a csontfejlődésben fontos szerepet játszó gén esetében a barkás agancs csontos részében sokkal nagyobb 10-szeres 30-szoros is lehet a génexpresszió, mint a vázcsontokban. Ezen gének közül a collA1 gén 1 és 5 kb-os promóter régiójában több runx2 transzkripciós faktor kötőhely található a gímszarvasban, mint az emberben vagy a szarvasmarhában, mert ezek a gének aktívabban működnek a gímszarvasban. A jövőben szeretnénk megvizsgálni, hogy vajon a többi csontfejlődéssel kapcsolatba hozható gén esetében is több runx2 vagy osx transzkripciós faktor kötőhely motívum található-e, vagy esetleg valamilyen SNP, vagy INDEL eltérés okozhat-e bármilyen osteogenezissel vagy osteoporozissal összefüggésbe hozható jelenséget. A runx2 és az osx transzkripciós faktorok

(mester regulátorok, központi szabályozó génjei a csontfejlődésnek) kötőhelyein kívül egyéb konzervált csontgén promóter motívumok is érdekesek lehetnek, tehát érdemes lenne többszörös szekvencia illesztéseket végezni közel rokon, vagy csontfejlődés szempontjából figyelemre méltó fajok és a gímszarvas ezen DNS szakaszai között. Az osteoporosis genomikai vizsgálatainak mintájára elvégezhetünk egyéb orvosi biológiai kutatásokat a szervfejlődés/regeneráció, robusztus szövet gyarapodás/tumor biológia témakörében.

6. ÚJ KUTATÁSI EREDMÉNYEK

A doktori kutatásom szervesen kapcsolódott a gímszarvas (*Cervus elaphus*) genom programhoz, amely a világ első elismert gímszarvas referencia genom összeállítása. Feladataim döntően a bioinformatika területére koncentráltak. A szerteágazó projekten belül a munkámhoz kapcsolódó új tudományos eredményeim az alábbiak:

- 1) A tudományos vizsgálatban használt kettős referencia vezérelt illesztéssel nagyobb pontossággal lehetett pszeudokromoszómákba rendezni a rövidebb gímszarvas DNS szekvenciákat (*de novo* scaffoldokat), mint az általánosan elterjedt egyetlen referenciára történő illesztés esetében. Referenciaként egy gímszarvas közeli rokon faj, a szarvasmarha (*Bos taurus*) jól meghatározott és annotált referencia genomja, továbbá a Slate-féle 34 kapcsoltsági csoportból és 621 marker pontból álló gímszarvas genetikai térkép szolgált. A marker pontokból azonban csak 361-et tudtam felhasználni az illesztésekhez.
- 2) A bioinformatikai munka során egy újonnan bevezetett algoritmus alapján dolgoztam (6. ábra.):
 - a) Megfeleltettem egymással a gímszarvas genetikai térkép markereket és a *de novo* gímszarvas scaffoldokat.
 - b) Megkerestem a gímszarvas géntérképi markerekhez nagyon hasonló szekvenciákat a szarvasmarha referencia genomon.
 - c) Azonosítottam a szarvasmarha referencia gének ortológ szekvenciáit a gímszarvas scaffoldokon.
 - d) Felillesztettem a *de novo* gímszarvas scaffoldokat az evolúciós változásoknak megfelelően átalakított szarvasmarha referencia genomra.
- 3) Minden több gént hordozó gímszarvas scaffold esetében látható, hogy az intra-scaffoldikus génjeik kiterjedt szinténiákat mutattak a szarvasmarha ortológ génekkel. Ebből következően a gének sorrendje nem csupán a térképpont

markerek által határolt szegmensekben és kromoszómális szinten, hanem a scaffoldokon belül is megegyezett.

- 4) A megfigyelt kapcsoltságok alapján elmondható, hogy a gímszarvas genom 81,5%-ában a gének az MMSc közötti szegmensekben a szarvasmarha ortológ gének sorrendjét, az MMSc-ken belül pedig a gímszarvas génsorrendet követik. A CerEla1.0 maradék 18,5%-ában, ahol a kromoszómális hasadások, fúziók és inverziók történtek a szarvasmarha és a gímszarvas génjeinek szinténikus blokkjai kombinálódtak.
- 5) Saját szkriptek készültek a különböző munkafázisokhoz.:
 - a) A BWA és MUMmer illesztőprogramok eredményfájljainak értelmezését és feldolgozását saját Bash szkriptekkel végeztem (20/a, 20/b, 20/c).
 - b) A meghatározott helyzetű scaffoldok kromoszómákká fűzéséhez Bash szkriptet írtam. A program a scaffoldok közötti rést az NCBI genomjainak mintájára 100 bázispárnyi N karakterrel töltötte fel (21/a, 21/b). (Végül azonban egy kollégám által készített Python szkript futtatásával létrehozott kromoszómákat töltöttük fel az NCBI-ba.)
- 6) A világon elsőként történt meg a gímszarvas gének szerkezetének és lokalizációjának meghatározása. Elkészült a kromoszómákba rendezett gímszarvas referencia genom teljes annotációja, beleértve ebbe a fehérje kódoló gének, a repetitív, a riboszómális RNS, a transzfer RNS és a mikro RNS szekvenciákat is.
- 7) A szarvasfélék kromoszómaínak sávtérképei és a szarvasmarha kromoszómák centromeron pozíciói alapján elhelyeztem a gímszarvas kromoszómák centromeronjait.
- 8) A bioinformatikai vizsgálat igazolta a gímszarvas kromoszómák feltételezett evolúciós változásait: 6 hasadás, 1 Robertsoniális transzlokáció és 1 Robertsoniális fúzió történt a szarvasmarha kromoszómákhoz képest. Bizonyítást nyert, hogy a Bt1 centromeron disztális karjával ekvivalens Ce19-ben lejátszódott egy hasadás és egy transzlokáció. A kromoszóma sávtérképek

és géntérképi markerek helyzete alapján a Ce28-ban valószínűleg egy paracentrikus inverziós esemény zajlott le.

9. Összességében bioinformatikai kutatómunkámmal érdemben hozzájárultam a világ első, nemzetközileg elismert kromoszómákba rendezett, teljes gímszarvas referencia genom összeállításához. A gímszarvas referencia genomot (CerEla1.0) feltöltöttük a szigorú szabályrendszerrel működő NCBI genomikai adatbázis és böngésző weboldalára. Így mindenki számára ingyenesen online elérhetővé, letölthetővé (MKHE00000000.1) tettük a teljes genom szekvenciát. Ennek köszönhetően 12 tudományos cikk született az elmúlt 3 évben, amelyben hivatkoznak a CerEla1.0 genomra.

7. AZ ÉRTEKEZÉS TÉMAKÖRÉBŐL ÍRT TUDOMÁNYOS KÖZLEMÉNYEK

Lektorált folyóiratban megjelent idegen nyelvű publikáció

Bana, N. Á., Nyiri, A., Nagy, J., Frank, K., Nagy, T., Stéger, V., Schiller, M., Lakatos, P., Sugár, L., Horn, P., Barta, E., & Orosz, L. (2018). The red deer *Cervus elaphus* genome CerEla1.0: sequencing, annotating, genes, and chromosomes. *Molecular genetics and genomics: MGG*, 293(3), 665–684. <https://doi.org/10.1007/s00438-017-1412-3>

Frank, K., Barta, E., Bana, N. Á., Nagy, J., Horn, P., Orosz, L., & Stéger, V. (2016). Complete mitochondrial genome sequence of a Hungarian red deer (*Cervus elaphus hippelaphus*) from high-throughput sequencing data and its phylogenetic position within the family Cervidae. *Acta biologica Hungarica*, 67(2), 133–147. <https://doi.org/10.1556/018.67.2016.2.2>

Lektorált folyóiratban megjelent magyar nyelvű publikáció

Bana, Á. N. (2020). A genom összerakás elmélete és alkalmazása a gímszarvas genom projektben. *Acta Agraria Kaposváriensis*, 24(1), 14-34. <https://doi.org/10.31914/aak.2370>

Nem lektorált folyóiratban megjelent magyar nyelvű publikáció

Orosz, L., & Bana, N. Á. (2019). Mire jó a szarvasgenom? *Élet és Tudomány*, 74(16), 498-500.

Konferenciakiadványban teljes terjedelemben idegen nyelven megjelent

Bana, Á. N., Nyiri, A., Nagy, J., Frank, K., Nagy, T., Stéger, V., Schiller, M., Lakatos P., Sugár L., Horn P., Barta E., & Orosz L. (2018, August 5-10). *The Red Deer Cervus elaphus Reference Genome CerEla1.0*. 9th International Deer Biology Congress, Estes Park, Colorado, USA, https://static.sched.com/hosted_files/idbc2018/3e/459317.pdf

Frank K., Barta, E., Bana, N.Á., Nagy, J., Horn, P., & Orosz, L., & Stéger, V., (2016, március 21-22). *Complete mitochondrial genome of the hungarian red deer (Cervus elaphus hippelaphus)*. Fiatal Biotechnológusok Országos Konferenciája 2016, Szent István Egyetem, Gödöllő, oldalszám: 69. ISBN 978-963-269-536-5.

Disszertáció témakörén kívüli: Lektorált folyóiratban megjelent idegen nyelvű publikáció

Frank, K., Bana, N. Á., Bleier, N., Sugár, L., Nagy, J., Wilhelm, J., & Stéger, V. (2020). Mining the red deer genome (CerEla1.0) to develop X-and Y-chromosome-linked STR markers. *Plos One*, 15(11). doi:10.1371/journal.pone.0242506

Előadások

Bana Á. N. (2019) A gímszarvas genom program bővítése (CerEla1.0) egy nagy sűrűségű SNP alapú géntérképpel. ÚNKP beszámoló, Kaposvár, 2019.04.03.

Bana Á. N (2018) A Gímszarvas/Csodaszarvas Genom Program, GENETIKAI MŰHELYEK MAGYARORSZÁGON” XVII. Minikonferencia, előadás. Szeged, 2018.09.21.

Bana Á. N., Nyiri A., Nagy J., Frank K., Nagy T., Stéger V., Schiller M., Lakatos P., Sugár L., Horn P., Barta E., Orosz L. (2018) The Red Deer *Cervus elaphus* Reference Genome CerEla1.0. 9th International Deer Biology Congress, Estes Park, Colorado, USA, August 5–10, 2018. pp. 163-164

Frank K., Barta E., Bana Á. N., Nagy J., Horn P., Orosz L., Stéger V. (2016) A magyarországi gímszarvas (*Cervus elaphus hippelaphus*) teljes mitokondriális genomja. F fiatal Biotechnológusok Országos Konferenciája „FIBOK 2016”, 2016. március 22.

Bana Á. N. (2016) Az első gímszarvas referencia genom szekvencia összerakása és annotálása. MBK napok 2016, Mezőgazdasági Biotechnológiai Kutatóintézet, Gödöllő, 2016. december 15.

Bana Á. N. (2016) A gímszarvas (*Cervus elaphus hippelaphus*) csont és agancs metabolizmus gének promóter szekvenciáinak bioinformatikai vizsgálata. F fiatal Biotechnológusok Országos Konferenciája 2016. Szent István Egyetem, Gödöllő, 2016. március 21.

Bana Á. N. (2015) Nagy teljesítményű genom analízisek bevezetése a déldunántúli gímszarvas populáció állományáiban, Kaposvár 2015. június 1.

Bana Á. N. (2013) A DeerPlex-ben felhasznált lókuszok kapcsoltsági vizsgálata, Gímszarvas workshop Bószénfa, 2013. június 25.