

Genotipizacija lokusa glavnog sustava tkivne podudarnosti srne, *Capreolus capreolus* (Linnaeus, 1758), sekvenciranjem nove generacije

Šurina, Mihaela

Master's thesis / Diplomski rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:548775>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-04-28**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)



Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Mihaela Šurina

**Genotipizacija lokusa glavnog sustava tkivne
podudarnosti srne, *Capreolus capreolus* (Linnaeus,
1758) sekvenciranjem nove generacije**

Diplomski rad

Zagreb, 2022.

University of Zagreb
Faculty of Science
Department of Biology

Mihaela Šurina

**Genotyping of the roe deer, *Capreolus capreolus*
(Linnaeus, 1758), major histocompatibility complex
by next generation sequencing**

Master thesis

Zagreb, 2022.

Ovaj rad je izrađen u Laboratoriju za molekularnu ekologiju na Fakultetu prirodoslovja, matematike i informacijske tehnologije Sveučilišta Primorska u Kopru i na Zavodu za animalnu fiziologiju Biološkog odsjeka Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu, pod voditeljstvom izv. prof. dr. sc. Ane Galov i prof. dr. sc. Elene Bužan. Rad je predan na ocjenu Biološkom odsjeku Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu radi stjecanja zvanja magistra molekularne biologije.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Diplomski rad

Genotipizacija lokusa glavnog sustava tkivne podudarnosti srne, *Capreolus capreolus* (Linnaeus, 1758) sekvenciranjem nove generacije

Mihaela Šurina

Rooseveltov trg 6, 10000 Zagreb, Hrvatska

Geni glavnog sustava tkivne podudarnosti (eng. *major histocompatibility complex*, MHC) jedni su od najvažnijih biomarkera za procjenu genetičke raznolikosti populacija jer su izrazito varijabilni. Zbog uloge receptora MHC u vezanju i prikazivanju antigena novonastale mutacije omogućuju otpornost jedinki na veći broj patogena i rezultiraju velikim brojem gena i alela. Probleme u genotipizaciji predstavljaju višestruka polimorfna mjesta u alelima te je uz tradicionalnu metodu sekvenciranja po Sangeru potrebno raditi predkorak molekularnog kloniranja kako bi se razdvojile alelne faze. Metode sekvenciranja nove generacije (eng. *next generation sequencing*, NGS) omogućuju zasebno, a istovremeno sekvenciranje velikog broja fragmenata DNA te uvelike pojednostavljaju i ubrzavaju sam proces. Primjenila sam jednu od metoda NGS, Ion Torrent S5, u genotipizaciji lokusa DRB sustava MHC srne (*Capreolus capreolus*) kao primjera nemodelnog kraljevnjaka. Umnožila sam egzon 2 lokusa DRB 14 uzoraka i napravila knjižnicu fragmenata za sekvenciranje. Dobivena očitavanja analizirala sam u programskom paketu AmpliSAT i identificirala sam osam alela od kojih je pet poznatih, a tri su novootkrivena. Budući da u srna postoji samo jedan lokus DRB te su očitavanja svakog uzorka bila izrazito velike pokrivenosti rezultati su bili jasni i jednoznačni. Primjenjivost tehnologije Ion Torrent potrebno je dalje testirati na uzorcima s udvostručenim lokusima te usporediti s rezultatima drugih metoda sekvenciranja.

(45 stranice, 17 Slika, 10 Tablica, 26 literaturnih navoda, jezik izvornika: hrvatski)

Rad je pohranjen u Središnjoj biološkoj knjižnici

Ključne riječi: NGS, MHC, Ion Torrent, AmpliSAT, priprema knjižnice

Voditelj 1: Izv. prof. dr. sc. Ana Galov

Voditelj 2: prof. dr. sc. Elena Bužan

Ocjenitelji:

Izv. prof. dr. sc. Ana Galov

Izv. prof. dr. sc. Maja Matulić

Izv. prof. dr. sc. Duje Lisičić

Rad prihvaćen: 13.1.2022.

BASIC DOCUMENTATION CARD

University of Zagreb
Faculty of Science
Department of Biology

Master Thesis

Genotyping of the roe deer, *Capreolus capreolus* (Linnaeus, 1758), major histocompatibility complex by next generation sequencing

Mihaela Šurina

Rooseveltova trg 6, 10000 Zagreb, Hrvatska

Genes of the major histocompatibility complex (MHC) are one of the most important biomarkers for assessment of genetic diversity of populations owing to their remarkable polymorphism. Due to role of MHC receptors in antigen binding and presentation, new mutations allow for resistance to a larger number of pathogens and result in a vast amount of genes and alleles. Multiple polymorphic sites in alleles, on the other hand, pose problems for successful genotyping by traditional methods and require application of molecular cloning in addition to Sanger sequencing in order to resolve allele phases. Next generation sequencing methods (NGS) have made possible sequencing of many DNA fragments individually, but in massive parallel fashion, thereby simplifying and speeding up the whole procedure. In this study I applied one of the NGS technologies, Ion Torrent S5, in genotyping MHC DRB locus of roe deer (*Capreolus capreolus*) as one of the nonmodel vertebrates. I amplified second exon of DRB locus in 14 individuals and constructed a sequencing library. I used online program AmpliSAT for data analysis and I identified eight alleles, five of which are already known and three that are new. Since roe deer has single DRB locus and sequencing yielded reads with great coverage, analysis was simple and straightforward. Applicability of Ion Torrent in genotyping needs to be further evaluated in species with duplicated DRB loci and in comparison with results of different sequencing technologies.

(45 pages, 17 figures, 10 tables, 26 references, original in: croatian)

Thesis is deposited in Central Biological Library.

Keywords: NGS, MHC, Ion Torrent, AmpliSAT, library preparation

Supervisor 1: Assoc. Prof. dr. sc. Ana Galov

Supervisor 2: Prof. dr. sc. Elena Bužan

Reviewers:

Assoc. Prof. dr. sc. Ana Galov

Assoc. Prof. dr. sc. Maja Matulić

Assoc. Prof. dr. sc. Duje Lisičić

Thesis accepted: 13.1.2022.

SADRŽAJ

| | |
|--|----|
| 1. UVOD | 1 |
| 1.1. Glavni sustav tkivne podudarnosti (MHC)..... | 1 |
| 1.1.1. Struktura i uloga proteina MHC | 1 |
| 1.1.2. Evolucija i organizacija gena MHC | 2 |
| 1.1.3. Važnost istraživanja gena MHC | 4 |
| 1.2. Metode genotipizacije MHC lokusa | 5 |
| 1.2.1. Sekvenciranje nove generacije..... | 6 |
| 1.2.1.1. Tehnologija Ion Torrent | 7 |
| 1.3. Srna (<i>Capreolus capreolus</i>)..... | 9 |
| 2. CILJ ISTRAŽIVANJA | 10 |
| 3. MATERIJALI I METODE | 11 |
| 3.1. Uzorci | 11 |
| 3.2. Mjerenje koncentracije DNA..... | 11 |
| 3.3. Lančana reakcija polimerazom (PCR)..... | 11 |
| 3.4. Elektroforeza u agaroznom gelu | 13 |
| 3.5. Priprema knjižnice fragmenata za sekvenciranje | 13 |
| 3.5.1. Spajanje replika produkata PCR i njihovo pročišćavanje | 13 |
| 3.5.2. Spajanje pročišćenih produkata PCR u zajedničku knjižnicu..... | 14 |
| 3.6. Provjera kvalitete pripremljene knjižnice na bioanalizatoru | 14 |
| 3.7. Priprema za sekvenciranje | 15 |
| 3.7.1. Emulzijska reakcija PCR | 16 |
| 3.7.2. Inicijalizacija Ion Torrent™ 5S sekvenatora | 17 |
| 3.7.3. Obogaćivanje čestica ISP | 17 |
| 3.7.4. Nanošenje uzorka na čip i pokretanje sekvenatora | 18 |
| 3.8. Analiza podataka | 19 |
| 3.8.1. Modul AmpliCLEAN | 20 |
| 3.8.2. Modul AmpliCHECK | 20 |
| 3.8.3. Modul AmpliSAS | 20 |
| 3.8.4. Ostali alati korišteni za analizu sekvenci | 21 |
| 4. REZULTATI..... | 22 |
| 4.1. Koncentracija izolirane DNA | 22 |

| | |
|---|----|
| 4.2. Produkti PCR dobiveni nakon umnažanja egzona 2 lokusa DRB | 22 |
| 4.3 Rezultati provjere kvalitete pripremljene knjižnice na bioanalizatoru | 22 |
| 4.4. Rezultati pripreme za sekvenciranje i sekvenciranja..... | 25 |
| 4.5. Analiza dobivenih očitavanja | 26 |
| 5. RASPRAVA..... | 34 |
| 6. ZAKLJUČAK | 39 |
| 7. LITERATURA..... | 40 |
| 8. PRILOZI..... | 44 |
| 9. ŽIVOTOPIS | 45 |

Popis kratica

μM - μmol/L

CIT – parametar postotak sličnosti (eng. *clustering identity treshold*)

DNA - deoksiribonukleinska kiselina

dNTP – deoksinukleotidni trofosfat

dsDNA – dvolančana DNA (ds- *double-stranded*)

ES – sustav za obogaćivanje (*enrichment system*)

FU – jedinice flouescencije, (*flouescence units*)

g - gram

GB - gigabajt

ISP - Ion Sphere™ particles,

mA – miliamper

MAF - parametar najmanja učestalost amplikona (eng. *minimum amplicon frequency*)

MB - megabajt

MDF - parametar najmanje učestalosti u odnosu na dominantnu sekvencu (eng. *minimum dominant frequency*)

MHC - glavni sustav tkivne podudarnosti (eng. *major histocompatibility complex*)

mL - mililitar

OT2 – Ion one touch 2

pb - parovi baza

PCR - lančana reakcija polimerazom (eng. *polymeraze chain reaction*)

pH - mjera kiselosti/lužnatosti ($\text{pH} = -\log [\text{H}^+]$)

pM – pikomolarna koncentracija, pmol/L

RFLP – polimorfizam duljine restrikcijskih fragmenata (eng. *restriction fragment lenght polymorphism*)

SAM - sequence alignment map

SSCP –polimorfnizam jednolančanih konformacija DNA (eng. *single-strand conformation polymorphism*)

TB - terabajt

TBE – tris - borat – EDTA

ubam - unmapped binary alignment map

UV – ultraljubičasta (eng. *ultraviolet*)

μL – mikrolitar

1.UVOD

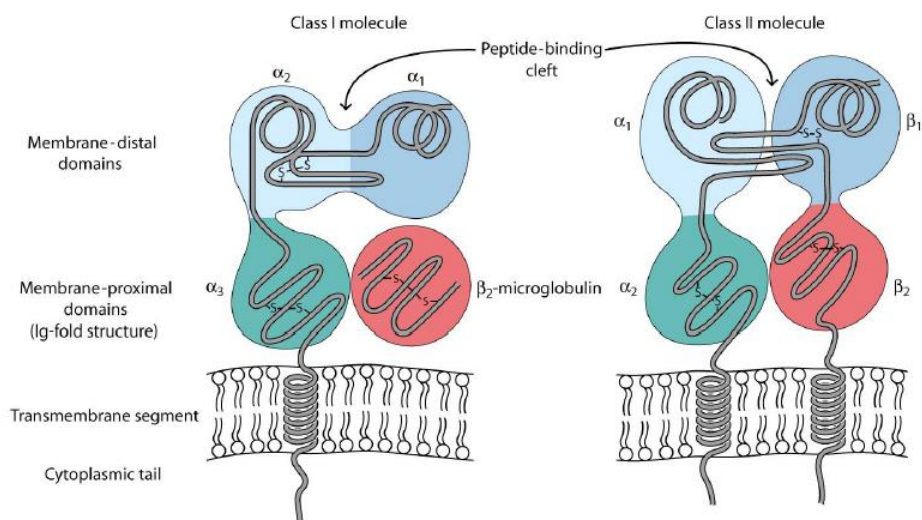
1.1. Glavni sustav tkivne podudarnosti (MHC)

1.1.1. Struktura i uloga proteina MHC

Proteini glavnog sustava tkivne podudarnosti (eng. *major histocompatibility complex*, MHC) su membranski receptori koji sudjeluju u imunološkim reakcijama posredovanim limfocitima T. Po strukturi i funkciji dijele se u tri skupine. Molekule skupine I i II sudjeluju u prezentaciji antigena, dok molekule skupine III sudjeluju u drugim imunološkim procesima i staničnoj signalizaciji (Kelley i sur., 2004).

Za razliku od limfocita B koji mogu neposredno prepoznati i vezati strane antigene putem svojih receptora, limfociti T mogu prepoznati samo antigene prikazane na MHC receptorima. MHC receptori skupine I nalaze se na svim stanicama s jezgrom i vežu antigene koji potječu iz unutrašnjosti stanice što mogu biti vlastiti stanični proteini ili proteini unutarstaničnih patogena kod zaraženih stanica. MHC receptore skupine II imaju samo stanice imunološkog sustava koje su specijalizirane za prikazivanje antigena (dendritičke stanice, makrofagi i limfociti B) i u procesu sklapanja oni se povezuju s antigenima koji potječu iz endosoma, odnosno iz izvanstanične tekućine. U slučaju prikazivanja stranih antigena spajanjem receptora MHC s receptorima limfocita T aktiviraju se imunološki odgovori kao što su upala, proizvodnja protutijela i uništenje zaraženih stanica izazivanjem apoptoze. Upravo zbog ovih procesa dolazi do odbacivanja transplatiranih tkiva po čemu su ovi proteini dobili ime (Kindt i sur., 2007; Alberts i sur., 2008).

Građa ovih proteina je slična (Slika 1). Molekule skupine I sastoje se od α polipeptidnog lanca čije distalne izvanstanične domene α_1 i α_2 tvore utor za vezanje antigena (vezno mjesto), a proksimalna α_3 domena interagira s receptorima limfocita T. Sadrže jednu transmembransku i citosolnu domenu, a α lanac je nekovalentno povezan s β_2 -mikroglobulinom kodiranim izvan MHC lokusa. Molekule skupine II sastoje se od α i β lanaca od kojih svaki sadrži dvije izvanstanične domene i po jednu transmembransku i citosolnu domenu. Vezno mjesto tvore domene α_1 i β_1 , a domena α_2 interagira s receptorima limfocita T (Alberts i sur., 2008).



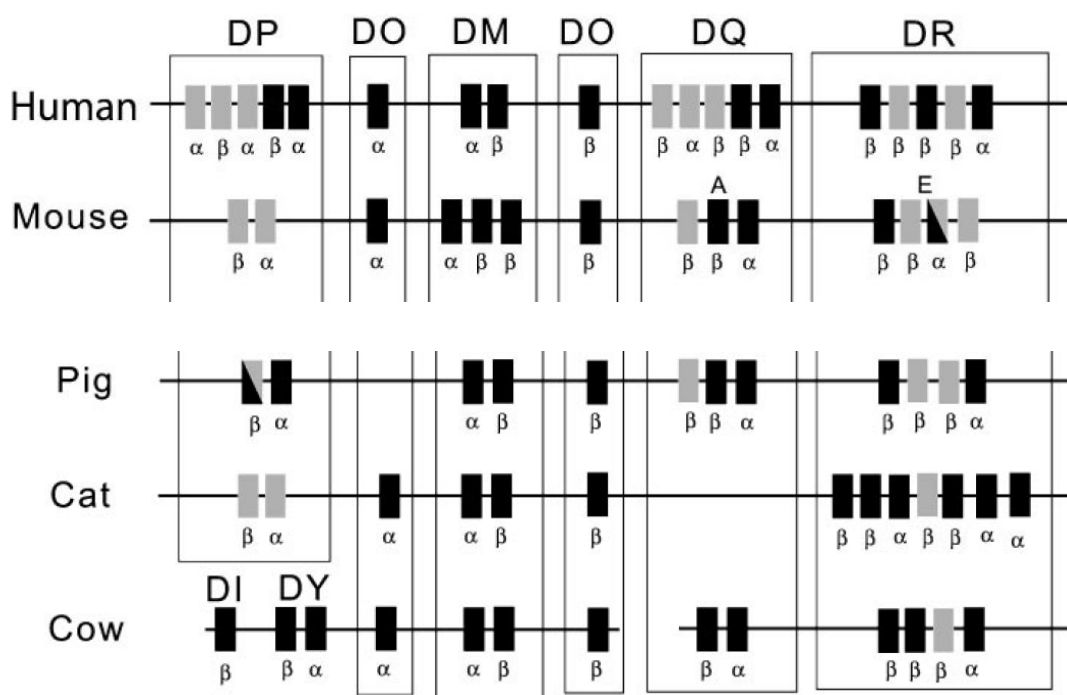
Slika 1. Građa proteina MHC skupine I i II (preuzeto iz Kindt i sur., 2007).

Molekule unutar iste skupine razlikuju se u varijabilnim izvanstaničnim domenama i prema tome u strukturi antigenih fragmenata koje mogu vezati. Koji će se fragment vezati na receptor MHC ovisi o njegovom pristajanju u vezno mjesto i koncentraciji u stanici ili izvanstaničnoj tekućini. Jedno vezno mjesto može vezati nekoliko različitih fragmenata antigena, ali samo jedan u određenom trenutku tako da stanice koje izražavaju veći broj istih ili različite receptore mogu reagirati na veći broj patogena (Kindt i sur., 2007). Zbog toga se mutacije gena MHC kroz evoluciju zadržavaju i rezultiraju velikom genetičkom raznolikošću (Ujvari i Belov, 2011).

1.1.2. Evolucija i organizacija gena MHC

Geni MHC pojavili su se kod prvih čeljustoustih kralježnjaka i od tada brzo evoluiraju kroz mnogostruke mutacije, što na razini gena, što na razini kromosoma, rezultirajući složenom organizacijom te velikim brojem gena i alela. Kod svih organizama geni MHC nalaze se u jednoj regiji velike genske gustoće, zajedno s nekim imunoregulacijskim i drugim genima, najčešće u poretku MHC I – MHC III – MHC II. Dok su ostali geni te regije i geni MHC skupine III znatno konzervirani, geni MHC skupine I i II pokazuju mali stupanj homologije. Evolucijska linija tih gena isprekidana je uslijed mnogih lokalnih duplikacija i inaktivacija zbog čega organizmi različitih razreda, čak i redova, imaju različite ancestralne gene te različit broj gena (Kelley i sur., 2004; Kumanovics i sur., 2003).

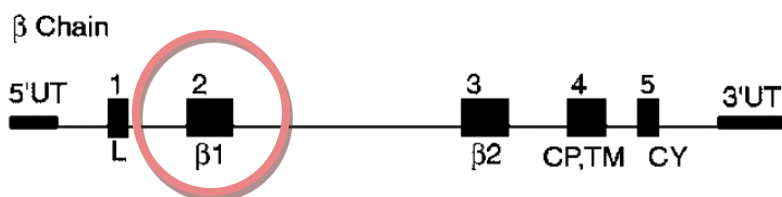
Geni MHC skupine II relativno su konzervirani kod sisavaca i homologija je prisutna na razini lokusa koji se razlikuju prema broju funkcionalnih gena i pseudogena (Slika 2). Najvarijabilniji su lokusi DR, DQ i DP čiji geni kodiraju za α i β lance molekula MHC skupine II, dok geni DM i DO kodiraju za pomoćničke proteine i nisu varijabilni (Mellins i Stern, 2014). Neke vrste su određene lokuse izgubile, a zadržale veći broj gena drugih lokusa (mačke) ili razvile analogne lokuse (goveda). Broj gena unutar lokusa može varirati čak i između jedinki; kod ljudi lokus DRB može imati od dva do šest funkcionalnih gena (Kumanovics i sur., 2003).



Slika 2. Shematski prikaz lokusa MHC skupine II kod nekoliko rodova sisavaca. Funkcionalni geni su prikazani crno, a pseudogeni sivo (preuzeto iz Kumanovics i sur., 2003).

Broj različitih molekula MHC koje prikazuje jedna jedinka kod većine vrsta nije velik; kod tipičnog heterozigotnog čovjeka to je 20-ak molekula, no raznolikost gena MHC na razini populacije veća je od svih drugih gena zajedno i jedinstvena po tome što se aleli međusobno razlikuju u velikom broju polimorfnih mjesta, a lokusi u broju gena (Van den Bussche i sur., 1999). Geni MHC skupine I generalno brže evoluiraju i polimorniji su. Homologija je očuvana tek na razini rodova i porodica i varijacije u broju gena veće su nego kod gena MHC skupine II što ih čini težim za proučavanje. Od gena MHC skupine II najpolimorniji je lokus DRB koji kodira za β lanac proteina (Slika 1), zbog čega je i najčešće analiziran u istraživanjima genetičke raznolikosti (Kumanovics i sur., 2003).

Najviše mutacija nalazi se u drugom egzonu (Slika 3) koji kodira domenu veznog mjesta i upravo ću njega sekvencirati u ovom istraživanju.



Slika 3. Struktura gena MHC skupine II kod čovjeka. Egzoni su prikazani kao pravokutnici s rednim brojem iznad, a kraticom regije koju kodiraju ispod. L – vodeći peptid (eng. leader peptide), β_1 i β_2 - izvanstanične domene, CP – vezni peptid (eng. connecting peptide); TM – transmembranska domena; CY, citoplazmatska domena, UT – netranslatirajuće regije (eng. untranslated regions) (preuzeto iz Bishop, 1998).

1.1.3. Važnost istraživanja gena MHC

Genetička raznolikost jedan je od najvažnijih pokazatelja fitnesa populacije jer omogućuje jedinkama da uspješno reagiraju na što više čimbenika, bolje se prilagode okolišnim uvjetima te donesu na svijet zdravo potomstvo. Zbog toga je izrazito bitno da se prije razvijanja konzervacijskih mjera, posebice za male i ugrožene populacije, procijeni genetička raznolikost kako bi se odredio prioritet i kako bi se pronašla odgovarajuća strategija (Perez i sur., 2019).

Većina čimbenika koji utječu na genetičku raznolikost mogu se svrstati u jednu od četiri kategorija: mutacije, koje su ishodište cjelokupne genetičke raznolikosti, protok gena, genetički drift - nasumične promjene u učestalosti alela i prirodnu selekciju. Za procjenu genetičke raznolikosti najčešće se koriste neutralni genetički markeri koji su dio nekodirajuće DNA i informiraju nas o prva tri čimbenika koji utječu na genetičku raznolikost. Kako ne kodiraju za funkcionalne produkte nisu pod utjecajem prirodne selekcije zbog čega se smatraju reprezentativnim primjerom genetičke raznolikosti genoma u cjelini i mogu dati uvid u evolucijske udaljenosti i srodstvene odnose unutar i između populacija (Ujvari i Belov, 2011). No, evolucija neutralnih markera često ne prati brze promjene u ekosustavu koje ostavljaju traga na genima pod utjecajem selekcije. Geni MHC su jedni od rijetkih genetičkih markera u molekularnoj ekologiji koji su dio kodirajuće DNA zbog izrazite varijabilnosti na razini populacije. Njihova raznolikost može biti sačuvana čak i kod populacija koje su prošle kroz drastične promjene zbog genetičkog drifta i zbog toga imaju opće nisku genetičku raznolikost mjerenu upotrebom neutralnih markera kao što su mikrosateliti i polimorfizam

duljine restrikcijskih fragmenata (eng. *restriction fragment length polymorphism*, RFLP) (Sommer, 2005).

Zbog svoje funkcije geni MHC su pod utjecajem ravnotežne selekcije koja održava novonastale alele jer heterozigoti imaju selekcijsku prednost pred homozigotima, zbog mogućnosti prepoznavanje većeg broja antigena i jer rijetki aleli s vremenom povećavaju učestalost u populaciji zbog čestih mutacija patogena (Perez i sur., 2019).

Proučavanjem gena MHC možemo dobiti uvid u interakcije jedinki s okolišem i međusobno te selekcijske pritiske koji djeluju na populaciju i njihovu adaptivnu sposobnost. Razvojem novih metoda sekvenciranja genotipizacija MHC lokusa postaje jednostavnija te se sve češće počinje upotrebljavati, uz neutralne genetičke markere, u procjeni genetičke raznolikosti populacija (Ujvari i Belov, 2011; Sommer, 2005).

1.2. Metode genotipizacije MHC lokusa

Otkrićem uloge proteina MHC i njihove važnosti za fitnes populacija 80-ih godina prošlog stoljeća počela su prva istraživanja varijabilnosti MHC gena kod divljih životinja. No, zbog složene organizacije, velike polimorfnosti i varijabilnog broja lokusa te mnoštva pseudogena još uvijek ne postoji metoda kojom bi genotipiziranje gena MHC bilo potpuno precizno i pouzdano. Neke od prvih metoda detekcije polimorfizama temeljile su se na promjeni konformacije DNA ovisno o sekvenci, no razlučivost takvih metoda drastično opada s povećanjem broja alela. Metoda promjene konformacije jednolančane DNA u nativnoj gel elektroforezi (eng. *single-strand conformation polymorphism*, SSCP) može razlučiti promjene u jednoj bazi i još uvijek se koristi za određivanje genetičke raznolikosti životinja s jednostavnim MHC sustavom i malim brojem lokusa. U tim slučajevima se također može koristiti umjesto kloniranja za razdvajanje gametnih faza prije Sangerovog sekvenciranja (Babik, 2010; Stipoljev i sur., 2020).

Samo Sangerovo sekvenciranje pokazalo se neučinkovito zbog problema fazne dvoznačnosti; zbog velikog broja polimorfnih mjesta ne može se odrediti koji je genotip majčinskog, a koji očevog alela, budući da se sve molekule DNA sekvenciraju zajedno. Zbog toga je potreban predkorak razdvajanja gametnih faza kako bi dobili sekvence pojedinačnih molekula. To se najčešće postiže kloniranjem produkata reakcije PCR u plazmidne vektore, transformacijom bakterijskih stanica te sekvenciranjem izolirane plazmidne DNA. Iako se ova metoda teoretski može koristiti za uzorke bilo koje razine složenosti primjenjiva je jedino u

slučaju manjeg broja uzoraka i alela. Budući da često nastaje velik broj artefakata, što zbog reakcije PCR, što zbog bakterijskog popravka krivo sparenih baza, za uklanjanje istih i pouzdano genotipiziranje poželjno je sekvencirati što više klonova. Za veliki broj uzoraka ili složene sustave MHC kloniranje i Sangerovo sekvenciranje postaje dugotrajno i skupocjeno (Babik, 2010; Stipoljev i sur, 2020). Ipak, zbog nenadmašive točnosti od 99.999% (Kumar i sur., 2019) ova metoda još uvijek predstavlja zlatni standard i koristi se za provjeru i potvrdu rezultata ostalih metoda genotipizacije (Babik, 2010).

1.2.1. Sekvenciranje nove generacije

Metode sekvenciranja nove generacije (eng. *next generation sequencing*, NGS) skup su različitih metoda koje su se razvile tijekom i nakon projekta ljudskog genoma (*Human genome project*, 1990.- 2003.) kako bi sekvenciranje postalo masovnije, brže, jednostavnije i jeftinije. Zajednička karakteristika im je visoka protočnost odnosno mogućnost sekvenciranja mnogo različitih fragmenata odjednom i stvaranje velike količine podataka u kratkom vremenu. Neki autori pod nazivom NGS podrazumijevaju samo one metode koje se temelje na umnažanju reakcijom PCR prije sekvenciranja, poznate još kao metode druge generacije, dok neki ubrajaju i metode sekvenciranja cijelih pojedinačnih molekula DNA poznate kao metode treće generacije. Iste, iako u teoriji obećavajuće, još uvijek nisu našle toliku primjenu kao metode druge generacije te ih ovdje neću opisivati (Paarek i sur, 2011).

Za razliku od Sangerovog sekvenciranja gdje je mnoštvo istih molekula u jednoj reakcijskoj smjesi, u metodama druge generacije svaki fragment ima svoje mjesto na pločici ili u jažici te se sekvenca očitava za svaki položaj što omogućuje da se sekvenciraju milijuni različitih fragmenata odjednom. Postoji nekoliko različitih proizvođača i metoda očitavanja sekvence, no svima im je zajednički korak pripreme knjižnice fragmenata za sekvenciranje (u nastavku samo „knjižnica“): DNA se u procesu pripreme knjižnice fragmentira do određene veličine, ovisno o metodi, i ligira sa sekvencama – adapterima koji služe kao mjesto prijanjanja početnice za povezivanje s nosačem i za sekvenciranje. Ukoliko se sekvenciraju isti fragmenti, npr. amplikoni iz različitih vrsta ili tretmana, vežu se i barkodovi – jedinstvene sekvence koje omogućuju razlikovanje pojedinih uzoraka. Nakon vezanja na nosač DNA se umnaža u procesu reakcije PCR kako bi se zbog mnoštva istih fragmenata signal ugradnje baze umnogostručio i mogao detektirati. Samo sekvenciranje se u većini slučajeva temelji na replikaciji DNA, kao i Sangerovo, a ono što se razlikuje su metode detekcije ugrađene baze (Paarek i sur, 2011; Kumar i sur., 2019).

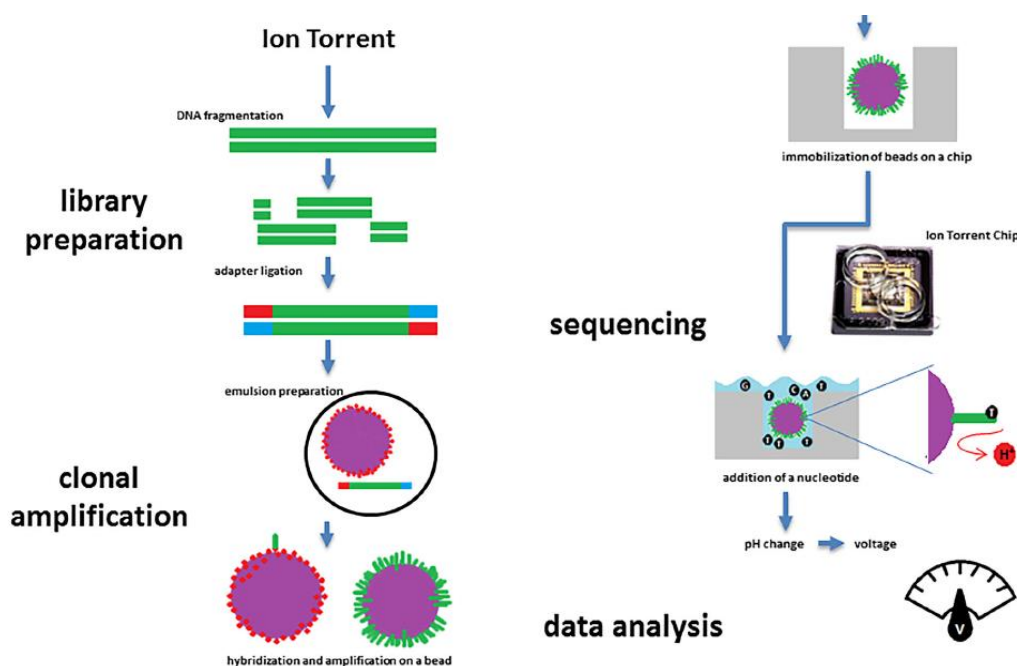
Najveće mane metoda NGS u usporedbi sa Sangerovom su zahtjevi za kratkim fragmenatima i manja točnost očitavanja sekvenci. Osim toga, zbog neophodne reakcije PCR prije sekvenciranja često dolazi do ugradnje krivih baza i stvaranja artefakata te može doći do promjene zastupljenosti pojedinih fragmenata zbog različite učinkovitosti umnažanja (Paarek i sur., 2011; Babik, 2010). To je posebice problematično kad proučavamo složeni sustav MHC s udvostručenim lokusima, no velika je prednost što možemo prilagoditi pokrivenost uzorka, odnosno koliko puta se određena baza „pročita“, prema potrebama istraživanja. Ako proučavamo polimorfizme visoko varijabilnih lokusa ili sekvenciramo *de novo* možemo pripremiti knjižnicu iz manjeg broja uzoraka i time dobiti veću pokrivenost svakog fragmenta te lakše raspoznavanje artefakata od pravih alela i pronalaženje rijetkih alela. Ovakva podesivost eksperimentalnih uvjeta i razvoj brojnih metoda i algoritma za obradu podataka pružaju mogućnost za prevladavanje poteškoća u genotipiziranju visoko varijabilnih i složenih sustava kao što je MHC (Babik, 2010; Stipoljev i sur, 2020).

1.2.1.1. Tehnologija Ion Torrent

Ion Torrent, kojeg ću koristiti u ovom istraživanju jedna je od novijih metoda NGS koja je stekla široku primjenu zbog jednostavnosti i brzine kojom očitava sekvence. Temelji se na mjerenju pH otopine tijekom replikacije DNA. Knjižnica fragmenata za sekvenciranje nanosi se na čip koji se sastoji od mnoštva malih jažica od kojih svaka funkcionira kao mini pH-metar (Slika 4). Ugradnjom baze otpušta se vodikov ion koji smanjuje pH otopine u jažici te se kemijski signal pretvara u električni. Otopine pojedinačnih nukleotida prelijevaju se preko čipa u ciklusima tako da dolazi do ugradnje samo jednog od četiri nukleotida u pojedinom ciklusu, na mjestima gdje se u lancu kalupu nalazi komplementarni nukleotid. U slučaju homopolimernih regija promjena pH se umnogostručuje za broj ugrađenih nukleotida (Paarek i sur, 2011; Kumar i sur, 2019).

Nakon pripreme knjižnice fragmenti se vežu za kuglice – ion sfere tj. čestice ISP (eng. *Ion Sphere™ particles*, ISP) i umnažaju u procesu emulzijske reakcije PCR u kojem kapljice vode raspršene po uljnoj fazi tvore mikroreaktore i fizički odvajaju fragmente koji se umnažaju. Idealno, u jednoj je kapljici jedan fragment koji se povezuje s kuglicom i umnaža pokrivajući njenu površinu. Ta se smjesa kasnije nanosi na čip i svaka kuglica zauzima mjesto u jednoj jažici. Mnoštvo istih fragmenata omogućuje stvaranje dovoljno jakog signala za detekciju (Glenn, 2011).

Za razliku od ostalih metoda NGS kao što su Illumina ili Roche 454, Ion Torrent ne zahtijeva skupe boje i reagense niti bilježenje signala kamerom što uvelike skraćuje vrijeme i trošak sekvenciranja te veličinu uređaja. Još jedna prednost je što se emulzijska reakcija PCR, za razliku od *bridge* reakcije PCR kojeg koristi Illumina, odvija u zasebnom uređaju i predstavlja dodatnu razinu kontrole čime ne ograničava uspješnost reakcije sekvenciranja. Ukoliko nešto pođe po krivu u emulzijskoj reakciji PCR, taj se korak može ponoviti prije nego se smjesa nanese na čip (Glenn, 2011). Najveće mane sustava Ion Torrent su ograničena količina podataka koju može proizvesti (maksimalno 25GB, nasuprot 3TB na „Illumina NovaSeq 6000“) te sklonost krivom očitavanju homopolimernih regija zbog čega nije pogodan za sekvenciranje cijelih genoma i *de novo* sekvenciranje, ali vrlo je praktičan za svakodnevnu kliničku uporabu i za ciljano sekvenciranje pojedinih regija ili gena: sekvenciranje amplikona (Kumar i sur., 2019).



Slika 4. Shematski prikaz protokola sekvenciranja tehnologijom Ion Torrent (preuzeto iz Ballard i sur., 2020).

1.3. Srna (*Capreolus capreolus*)

Srne pripadaju porodici jelena (*Cervidae*) čiji su pripadnici ekološki važne vrste zbog toga što su hrana velikim zvijerima, a svojom ishranom kontroliraju biljnu vegetaciju i održavaju pašnjake. U proteklih 50-ak godina značajno su proširili svoja staništa na sjevernoj hemisferi prelazeći iz šuma na poljoprivredne površine. Veća količina i dostupnost hrane omogućila im je povećanje brojnosti, no smatra se da su zbog srodnosti s govedima pod utjecajem patogena domaćih životinja. Osim toga, postoji mogućnost da su u procesu kolonizacije novih staništa prošle kroz učinke utemeljitelja (eng. *founder effect*) što se može negativno odraziti na njihovu genetičku raznolikost i sposobnost obrane od novih patogena. Zbog toga su krajem 90-ih godina počela istraživanja varijabilnosti gena MHC na parnoprstašima, no mali broj njih koristi novije metode sekvenciranja (Quemere i sur., 2015).

U dosadašnjim istraživanjima (Mikko i sur., 1999; Van Den Bussche i sur., 1999; Quemere i sur., 2015) utvrđeno je da u srna lokus DRB nije udvostručen te jedinke imaju samo dva alela što uvelike pojednostavljuje proces analize podataka i razlikovanja stvarnih alela od artefakata. Zbog toga je srna jednostavan primjer nemodelnog kralježnjaka koji može poslužiti za prvobitnu usporedbu različitih metoda genotipizacije i pronalaženja one na kojoj bi se mogla temeljiti buduća istraživanja genetičke raznolikosti.

2. CILJ ISTRAŽIVANJA

Cilj ovog istraživanja je primijeniti metodu sekvenciranja nove generacije (NGS) u genotipizaciji nemodelnih kralježnjaka, konkretno na primjeru genotipizacije lokusa DRB u srne, te ispitati mogućnost optimizacije metode analize dobivenih rezultata. Rezultati ovog istraživanja ubuduće će se koristiti u svrhu usporedbe s ostalim metodama genotipizacije kako bi se pronašla ona koja daje najpouzdanije rezultate uz što jednostavniji protokol i koja će se primjenjivati u budućim imunogenetičkim istraživanjima različitih vrsta divljih životinja. Uz to, mogući novootkriveni aleli poslužit će u daljnjim istraživanjima procjene adaptivne raznolikosti srna.

3. MATERIJALI I METODE

3.1. Uzorci

Za ovo istraživanje koristila sam 15 uzoraka izolirane DNA srna. Uzorci su skupljeni u sklopu projekta Hrvatske zaklade za znanost „Interakcija nositelj-parazit: odnos tri različita tipa nositelja prema invaziji metiljem *Fascioloides magna*“ (IP-01-2018) voditelja izv. prof. dr. sc. Deana Konjevića s Veterinarskog fakulteta Sveučilišta u Zagrebu. DNA je izolirana na Zavodu za animalnu fiziologiju PMF-a komercijalnim kompletom „Wizard genomic DNA purification“ (Promega) i otopljena je u vodi bez nukleaza.

Odabrane su heterozigotne jedinke kojima se nije moglo jasno odrediti alele Sangerovim sekvenciranjem tako da ukupan uzorak u ovom istraživanju nije nasumičan. Jedinke potječu iz različitih hrvatskih populacija i označene su sljedećim oznakama; 1SN, 1SL, 2SL, 3SL, L2, L5, L19, L20, K7, K10, 4SC, 12SC, 13SC, 14SC, M8/35.

3.2. Mjerenje koncentracije DNA

Uzorcima izolirane DNA izmjerila sam koncentraciju Qubit™ 3.0 fluorometrom koristeći komercijalni komplet „Qubit™ dsDNA BR assay“ (ThermoFisher Scientific). Radnu otopinu pufera i boje pripremila sam tako da po uzorku sadrži 199 µL pufera i 1 µL boje. Za kalibraciju uređaja koristila sam 190 µL radne otopine i 10 µL standarda iz komercijalnog kompleta, a za mjerenje koncentracije uzoraka 198 µL radne otopine i 2 µL DNA. Uzorak M8/35 je imao koncentraciju ispod razine detekcije te smo ga isključili iz istraživanja.

3.3. Lančana reakcija polimerazom (PCR)

Na uzorcima izolirane DNA provela sam lančanu reakciju polimerazom (eng. *polymerase chain reaction*, PCR) kako bih umnožila egzon 2 lokusa DRB te istovremeno na njega vezala adaptere i barkodove. Koristila sam uzvodnu (eng. *forward*, F) početnicu LA31 i nizvodnu (eng. *reverse*, R) početnicu LA32 (Sigurdardottir i sur., 1991). Početnice su prethodno pripremljene na način da su na njih ligirani adapteri i barkodovi iz komercijalnog kompleta „Ion Xpress™ Barcode Adapters 1-96“ (ThermoFisher Scientific). F početnica na 5' kraju sadrži adapter A, barkod i vezni odsječak, dok R početnica sadrži samo adapter P1, također na 5' kraju. Adapter P1 služi za povezivanje fragmenata s česticama ISP u emulzijskoj reakciji PCR - na njega se veže početnica koja se nalazi na česticama ISP, a adapter A služi u reakciji sekvenciranja; na njega se veže početnica za sekvenciranje. Raspored nukleotidnih motiva početnica prikazan je na Slici 5. Početnice sam prije korištenja razrijedila na 10 µM (µmol/L).

5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGAGCCTATTCGATGATCCTCTCTCTGCAGCACATTTCTCT-

5'-CCACTACGCCTCCGCTTTCTCTCTATGGGCAGTCGGTGATTTCGCGTCACCTCGCCGCTG-

Slika 5. Shematski prikaz početnica s dodanim nukleotidnim motivima: F početnica obojana je žuto, a R ljubičasto. Plavo su označeni adapteri, crveno barkod, zeleno vezni odsječak, a crno sekvenca LA31 odnosno LA32.

Reakcijske smjese pripremala sam u laminaru, nakon sterilizacije pribora UV svjetlom, koristeći reagense iz komercijalnog kompleta „AllTaq PCR Core“ (Quiagen). Uzorke sam pripremila u triplikatima prema podacima iz Tablice 1. Za svaki uzorak napravila sam zasebnu matičnu otopinu s jedinstvenom F početnicom i razdijelila je u tri replike i jednu negativnu kontrolu. Kao negativnu kontrolu koristila sam vodu iz komercijalnog kompleta. Reakciju PCR sam provela u uređaju SimpliAmp (Applied Biosystems) prema temperaturnom programu iz Tablice 2.

Tablica 1. Komponente reakcijskih smjesa za PCR i njihovi volumeni.

| reagens | volumen u pojedinačnoj reakciji (μL) | volumen u matičnoj otopini (x4)(μL) |
|-------------------|--------------------------------------|-------------------------------------|
| Voda | 7,1 | 28,4 |
| 5x Pufer | 5 | 20 |
| MgCl ₂ | 2,5 | 10 |
| 5x Q otopina | 1 | 4 |
| dNTP | 0,5 | 2 |
| F početnica | 0,3 | 1,2 |
| R početnica | 0,3 | 1,2 |
| Taq polimeraza | 0,3 | 1,2 |
| DNA | 3 | / |
| <i>Ukupno</i> | 20 | 68 |

Tablica 2. Temperaturni program reakcije PCR.

| korak | temperatura (C°) | vrijeme | broj ponavljanja |
|----------------------|------------------|---------|------------------|
| početno zagrijavanje | 95 | 3 min | 1 |
| denaturacija | 95 | 30 s | 35 |
| vezanje početnica | 60 | 1 min | |
| produljenje | 72 | 1 min | |
| završno produljenje | 72 | 10 min | 1 |

3.4. Elektroforeza u agaroznom gelu

Za provjeru uspješnosti reakcije PCR napravila sam gel od 130 mL 0.5x TBE (Tris/Borat/EDTA) pufera i 1.9 g agaroze. Za vizualizaciju DNA koristila sam 8 µL boje *Midori Green* koju sam dodala u gel prije izlijevanja. Zbog osjetljivosti boje na svjetlo gel sam prekrila folijom za vrijeme polimerizacije kao i za vrijeme elektroforeze. Na komadu parafilma pomiješala sam 2 µL svake replike produkta PCR s podjednakom količinom 1x pufera za nanošenje „10x DNA Gel Loading Buffer“ (Omega Bio-Tek) i nanijela na gel. Kao pozitivnu kontrolu i marker molekularnih veličina koristila sam „GeneRuler 50 bp DNA Ladder“ (ThermoFisher Scientific). Elektroforeza je tekla 30 min pri 120 V i 400 mA. Rezultate sam zabilježila kamerom sustava za dokumentiranje gelova „MiniBis“ (DNR Bio-imaging Systems).

3.5. Priprema knjižnice fragmenata za sekvenciranje

Budući da egzon 2 lokusa DRB prema veličini zadovoljava kriterije knjižnice za sekvenciranje tehnologijom Ion Torrent nije bilo potrebno fragmentiranje ni ligacija sekvenci adaptera već su oni dodani u procesu reakcije PCR, stoga PCR možemo smatrati prvim korakom pripreme knjižnice. Ostali koraci pripreme knjižnice u ovom istraživanju bili su spajanje replika produkata PCR za svaki uzorak, pročišćavanje i mjerenje koncentracije spojenih replika produkata PCR, spajanje spojenih replika produkata PCR svakog uzorka u zajedničku knjižnicu tako da im koncentracije budu približno jednake, i ponovno pročišćavanje knjižnice.

Osim uzoraka korištenih za moj diplomski rad knjižnica je sadržavala još 82 uzorka kod kojeg je umnožen egzon 2 lokusa DRB s istim početnicama. Cilj je staviti što više uzoraka na čip kako bi sekvenciranje bilo isplativije, a u ovom slučaju ograničavajući čimbenik bio je broj barkodova, stoga je konačni broj različitih uzoraka koje sam pripremila za sekvenciranje bio 96.

3.5.1. Spajanje replika produkata PCR i njihovo pročišćavanje

Sve replike jednog uzorka spojila sam u mikroepruvete LoBind™ (Eppendorf) i provela pročišćavanje produkata PCR magnetnim kuglicama „Agencourt AMPure XP“ (Agencourt Bioscience Corporation) prema protokolu prilagođenom za duljinu fragmenata od ~400 pb. Magnetne kuglice imaju afinitet za fragmente DNA veće od 100 pb i potrebno je prilagoditi vrijeme inkubacije s kuglicama i ispiranja etanolom kako bi odstranili sve kratke

fragmente i neiskorištene reagense, a zadržali što više umnoženih fragmenata. Kuglice sam dodala u produkte PCR u količini 70% volumena produkta PCR, vorteksirala i kratko centrifugirala te inkubirala 5 min na sobnoj temperaturi kako bi se produkti PCR vezali na kuglice. Zatim sam ih stavila na stalak s magnetom i inkubirala 3 min kako bi se kuglice izdvojile iz otopine. Otpipetirala sam supernatant i isprala talog dva puta s 500 μL 70% etanola. Trajanje inkubacije s etanolom bilo je 30 s i za to vrijeme sam okretala mikroeprovete oko osi kako bi kuglice prošle kroz otopinu. Nakon kratkog sušenja (do 5 min) kako bi ispario preostali etanol maknula sam mikroeprovete s magnetnog stalka, dodala 30 μL vode bez nukleaza i vorteksirala 10 sekundi kako bi se DNA odvojila s kuglica. Ponovno sam stavila mikroeprovete na stalak i kad su se magneti odvojili, supernatant s pročišćenim produktima PCR prebacila sam u nove mikroeprovete.

3.5.2. Spajanje pročišćenih produkata PCR u zajedničku knjižnicu

Pročišćenim produktima PCR izmjerila sam koncentraciju prema postupku opisanom u poglavlju 3.2. i sve ih spojila u jednu mikroeprovetu - knjižnicu tako da budu približno jednakih koncentracija; prema Tablici 3. Nakon spajanja produkata PCR knjižnicu sam još jednom pročistila Agencourt magnetnim kuglicama prema ranije opisanom protokolu.

Tablica 3. Volumeni produkata PCR korišteni za izradu knjižnice.

| koncentracija produkta PCR (ng/μL) | volumen u knjižnici (μL) |
|---|---|
| <10 | 2 |
| 10-20 | 1 |
| 20-35 | 0.8 |
| <35 | 0.5 |

3.6. Provjera kvalitete pripremljene knjižnice na bioanalizatoru

Kako bih provjerila kvalitetu pročišćene knjižnice i preciznije odredila koncentraciju (u pmol/L, pM) pripremila sam 10x, 20x i 50x razrjeđenja knjižnice. Koristila sam uređaj „2100 Bioanalyzer Instrument“ s pripadajućim komercijalnim kompletom „Agilent high sensitivity DNA“ (Agilent Technologies). Komercijalni komplet sadrži reagense: gel matriks, boju, standarde, molekularni marker veličina te čipove koji sadrže jažice i kanaliće u kojima se odvija mini elektroforeza u gelu. Nakon što su se reagensi otopili zajedno s asistenticom pripremila sam čip prema uputama proizvođača. Ukratko, nanijeli smo smjesu gela i boje prvo u posebno označenu jažicu, nakon čega smo u aparaturi za pripremu čipa (eng. *chip priming*

station) razvukli gel, a zatim u preostale tri jažice u zadnjem stupcu (Slika 6). U svih 12 jažica dodali smo standarde koji služe kao unutarnje kontrole i koji predstavljaju najveću i najmanju veličinu fragmenta koja se može analizirati ovim komercijalnim kompletom. Zatim smo u jažicu označenu ljestvama dodali molekularni marker veličina, a od preostalih 11 jažica u tri smo nanijeli naša razrjeđenja knjižnice, a u ostale ponovno standarde (ukupno se može nanijeti 11 uzoraka, a mi smo imali tri). Čip smo vorteksirali u vorteksu za čipove *IKA MS3* (Agilent Technologies) i umetnuli ga u bioanalizator odmah nakon pripreme kako uzorci ne bi isparili. Uređaj smo pokrenuli preko računalnog programa „2100 Expert Software“ u kojeg sam ranije unijela odgovarajuće podatke o uzorcima i čipu te izašli iz prostorije kako bi uklonili sve izvore vibracija koji bi mogli omesti rezultate.



Slika 6. DNA čip visoke osjetljivosti („High Sensitivity DNA Chip“, Agilent Technologies). Slovom G i sivom pozadinom su označene jažice u zadnjem stupcu u koje se stavlja gel, ljestvama je označena jažica u koju se stavlja marker molekularnih veličina, a brojevima i plavom pozadinom jažice u koje se mogu staviti uzorci za analizu.

Razrjeđenje kod kojeg je bio najveći postotak fragmenata željene duljine uzeli smo za daljnju analizu i razrijedili do koncentracije od 100 pM.

3.7. Priprema za sekvenciranje

Nakon provjere kvalitete i preciznog određivanja koncentracije knjižnice te njenog razrjeđenja na željenu koncentraciju slijedilo je vezanje fragmenata na nosače – čestice ISP i umnažanje emulzijskom reakcijom PCR, obogaćivanje čestica ISP te nanošenje na čip. Koristila sam sustav „Ion One Touch 2“ (ThermoFisher Scientific) koji se sastoji od dva uređaja. Uređaj „Ion One Touch“ (Slika 7) provodi emulzijsku reakciju PCR, a „Ion One Touch ES“ (eng. *enrichment system*) obogaćivanje čestica ISP – izolaciju čestica na koje su vezani fragmenti (tzv. pozitivne čestice ISP), koja se provodi kako one na koje nisu vezani

fragmenti ne bi remetile proces sekvenciranja. Koristila sam komercijalni komplet „Ion 520™& Ion 530™ – OT2“ za duljinu fragmenata od 400 pb i za sve korake pratila upute proizvođača (Ion 520™& Ion 530™– OT2 korisnički upute). Upute se sastoje od mnoštva koraka i detalja koje ovdje neću navoditi, već ću ih samo ukratko opisati. Vrlo je bitno proučiti cijeli protokol prije početka pripreme jer su određeni koraci vremenski ograničeni i postoji samo nekoliko njih na kojima se možemo zaustaviti.

3.7.1. Emulzijska reakcija PCR

Protokol za emulzijsku reakciju PCR sastoji se od podešavanja uređaja, pripreme amplifikacijske otopine i sakupljanja pozitivnih čestica ISP nakon umnažanja. Najvažniji dijelovi uređaja Ion One Touch prikazani su na Slici 7.



Slika 7. Uređaj *Ion One Touch* i njegovi najvažniji dijelovi: 1. centrifuga 2. mjesto gdje se umeće amplifikacijski filter u kojem se stvaraju emulzijske kapljice 3. mjesto gdje se umeće amplifikacijska ploča koja omogućuje ravnomjerno zagrijavanje i hlađenje.

Mikroeprovete za sakupljanje čestica ISP (eng. *recovery tubes*) napunila sam otopinom za razbijanje emulzije i umetnula u centrifugu. Vanjske epruvete sam dopunila s emulzijskim uljem i otopinom za sakupljanje čestica ISP i zajedno s asistenticom umetnula amplifikacijsku ploču te spojila sve potrebne dijelove u uređaju. Amplifikacijsku otopinu pripremali smo u odvojenoj prostoriji, za provođenje procedura poslije reakcije PCR (eng. *post-PCR*). Napravili smo dodatno razrjeđenje knjižnice; 8 μ L knjižnice koncentracije 100 pM razrijedili smo do 100 μ L s vodom bez nukleaza. U mikroeprevetu koja već sadrži mješavinu reagensa za PCR dodali smo enzime, čestice ISP i tako razrijedenu knjižnicu. U

amplifikacijski filtar prvo smo nanijeli amplifikacijsku otopinu i zatim reakcijsko ulje te smo ga umetnuli u uređaj i pokrenuli program. Vrlo je bitno da se uređaj pokrene unutar 15 minuta od pripreme amplifikacijske otopine.

Trajanje emulzijske reakcije PCR je oko šest sati, a unutar 16 sati nakon njega uzorke je potrebno sekvencirati zbog čega smo pripremu napravili popodne i ostavili ga da radi preko noći kako bismo sljedeći dan napravili izolaciju i obogaćivanje čestica ISP i odmah nakon toga pokrenuli sekvenciranje.

Nakon završene emulzijske reakcije PCR uređaj je centrifugirao uzorke te sam otpipetirala dvije trećine supernatanta, a taloge resuspendirala u preostaloj tekućini. Resuspendirane taloge spojila sam u novu mikroeprijetu LoBind™. Mikroeprijetu za sakupljanje sam isprala s vodom bez nukleaze kako bi pokupila preostale čestice ISP i dodala u mikroeprijetu. Sve sam vorteksirala i ponovo centrifugirala, bez provjere kvalitete, te sam talog u konačnici resuspendirala u otopini za resuspenziju čestica ISP. U ovom trenutku se u mikroeprijeti nalaze čestice ISP na koje su vezani naši fragmenti, ali i prazne čestice ISP.

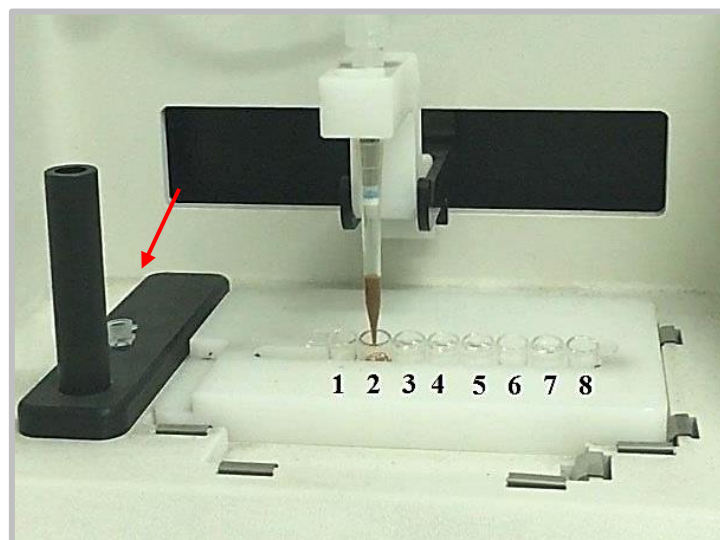
3.7.2. Inicijalizacija Ion Torrent™ S5 sekvencatora

Prije obogaćivanja čestica ISP na uređaju ES pokrenuli smo inicijalizaciju Ion S5™ sekvencatora (ThermoFisher Scientific) kako bi bili sigurni da je uređaj uspješno inicijaliziran prije nego što stavimo uzorak na čip koji se odmah nakon punjenja mora sekvencirati. U procesu inicijalizacije sekvencatora stare reagense zamjenjujemo novima i uređaj se priprema za početak sekvenciranja: ispraznili smo kontejner za otpad, stavili novu bocu otopine za ispiranje i novu kazetu s reagensima te provjerili da je čip od prethodnog sekvenciranja unutra jer je i on potreban za uspješnu inicijalizaciju.

3.7.3. Obogaćivanje čestica ISP

Za vrijeme inicijalizacije sekvencatora pripremila sam reagense za uređaj Ion One Touch ES. Potrebno je svaki put pripremiti svjež u otopinu za rastapanje (eng. *melt-off solution*) koja se sastoji od detergenta Tween i NaOH. Magnetne kuglice „Dynabeads™ MyOne™ Streptavidin C1“ (ThermoFisher Scientific) sam pripremila na način da sam 100 µL resuspendiranih kuglica stavila na magnetni stalak, talog isprala otopinom za ispiranje i resuspendirala u otopini za hvatanje kuglica (eng. *beads capture solution*). Trakicu s 8 jažica (Slika 8) napunila sam redom; uzorkom knjižnice (1), otopinom magnetnih kuglica (2), otopinom za ispiranje(3-5) i otopinom za rastapanje (7). U uređaj sam stavila i novu

mikroepruvetu za PCR u kojoj će se skupiti pročišćeni produkt i novi filter nastavak za pipetu od 1 mL. Nakon što je uređaj završio s radom (~35 min) provjerila sam volumen u mikroepruveti za PCR; ako je postupak uspio u mikroepruveti se nalazi barem dvije trećine volumena. Bitno je odmah isprati uzorak jer otopina za rastapanje koja se nalazi u mikroepruveti može oštetiti DNA. Mikroepruvetu sam centrifugirala, uklonila supernatant i talog u konačnici resuspendirala u vodi bez nukleaza. U ovom trenutku u mikroepruveti se nalaze samo čestice ISP s fragmentima za sekvenciranje – pozitivne čestice ISP.



Slika 8. Dio uređaja *Ion One Touch ES* u kojem se zbiva reakcija obogaćivanja čestica ISP; brojevima su označene jašice s reagensima, a strelicom mikroepruveta za PCR u kojoj se skuplja obogaćeni produkt.

3.7.4. Nanošenje uzorka na čip i pokretanje sekvenatora

Za vrijeme rada uređaja ES u računalnom programu „Torrent Suite software 5.12.1.“ unijela sam podatke o uzorcima, barkodovima, korištenim komercijalnim kompletima, uređajima i vrsti čipa. Nakon sekvenciranja rezultati su dostupni za preuzimanje u ovom programu.

Budući da je inicijalizacija sekvenatora bila uspješno završena odmah smo pripremili čip. Prije nanošenja uzorka na čip potrebno je na adapter vezati početnicu za sekvenciranje te dodati polimerazu. U mikroepruvetu za PCR s obogaćenim česticama ISP dodala sam pufer za vezanje početnice (eng. *annealing buffer*) i početnicu za sekvenciranje te kontrolne čestice ISP poznate sekvence koje služe za provjeru uspješnosti reakcije i određivanje kvalitete očitavanja. Navedeno sam stavila u uređaj za PCR na temperaturni program 2 min na 95°C i 2 min na 37°C kako bi se početnica za sekvenciranje vezala na adaptere. Nakon toga sam u

uzorak dodala polimerazu i pufer za nanošenje te ga inkubirala na sobnoj temperaturi. Koristili smo „Ion™ 530“ čip. Uzorci i ostali reagensi se na čip nanose kroz otvor za nanošenje, a otpipetiraju iz izlaznog bazenčića (Slika 9). Asistentica je nanijela pripremljeni uzorak na čip i postavila ga na centrifugu „Ion Chip™ MiniFuge“ (ThermoFisher Scientific) kako bi se čestice ISP potisnule u jažice na čipu. Za vrijeme centrifugiranja pripremili smo otopinu za pjenjenje (eng. *foaming solution*) koju je potom nanijela i odpipetirala s čipa, četiri puta. Pjena uzrokuje da se čestice ISP u jažicama poslože u jednom sloju. Zatim je još četiri puta isprala čip puferom za vezanje početnice i umetnula ga u sekvenator. Na zaslonu sekvenatora odabrali smo pripremljeni program i pokrenuli sekvenciranje. Prostorija sa sekvenatorima se za vrijeme rada zatvara i nije dozvoljeno ulaziti u nju zbog vibracija koje bi mogle omesti proces sekvenciranja (usmeno priopćenje).



Slika 9. Ion 530™ čip; Brojevi označavaju otvor (1) i bazenčić za nanošenje (2) te izlazni otvor (3) i izlazni bazenčić (4) (preuzeto sa

<https://www.thermofisher.com/order/catalog/product/A27763#/A27763>)

3.8. Analiza podataka

Za analizu podataka koristila sam *online* programski paket AmpliSAT koji je namijenjen za obradu podataka dobivenih sekvenciranjem amplikona, odnosno produkata PCR. Sastoji se od nekoliko modula putem kojih se sirova očitavanja grupiraju i uklanjaju prema zadanim parametrima kako bi se razdvojili pravi aleli od artefakata koji nastaju u procesu reakcije PCR (Sebastian i sur., 2016).

Sirova očitavanja sekvenci preuzela sam s programa Torrent Suite u obliku datoteka ubam (eng. *unmapped binary alignment map*). Takva vrsta datoteke sadrži sažete sekvence u binarnom kodu i prikladna je za skladištenje i prijenos velike količine podataka, ali je u ovom obliku nečitljiva. Kako je za analizu u programskom paketu AmpliSAT potrebno imati

datoteke fastq koje sadrže sekvence u pisanom obliku zajedno s pripadajućom ocjenom kvalitete (eng. *quality score*, q) pretvorbu sam izvršila preko programa SAMtools (eng. *sequence alignment map*, SAM) koji sadrži razne alate za uređivanje i analizu podataka dobivenih sekvenciranjem NGS (Li i sur., 2009).

3.8.1. Modul AmpliCLEAN

U modul AmpliCLEAN unijela sam dobivene dokumente fastq koje sam sažela u datoteke veličine do 500 MB. izradila sam datoteku s podacima o produktu PCR (eng. *amplicon data*) u kojoj sam prema uputama navela ime gena, očekivanu duljinu, sekvence početnica i ime vrste. Parametar „minimalne kvalitete očitavanja“ (eng. *minimum phred quality score*) nisam mijenjala, budući da su očitavanja dobre kvalitete i ima ih puno namjestila sam ga na 30 čime su se odstranile sekvence čija je točnost očitavanja manja od 99.9%. Najmanju duljinu sekvence stavila sam na 250, a najveću na 300 pb. Ove datoteke sam dalje koristila u modulima AmpliCHECK i AmpliSAS.

3.8.2. Modul AmpliCHECK

Modul AmpliCHECK služi za preliminarni pregled podataka prije detaljne analize u modulu AmpliSAS kako bi što bolje prilagodili parametre (Sebastian i sur., 2016).

Koristila sam zadane parametre za tehnologiju sekvenciranja Ion Torrent; 0,5% za supstitucijske pogreške, 1% za insercije/delecije (indel) i 1% za minimalnu učestalost amplikona (eng. *minimum amplicon frequency*). Maksimalni broj očitavanja koje program može obraditi je 5 000. Osim datoteke s općim podacima o amplikonu napravila sam i datoteku svih poznatih alela srna koje sam preuzela iz GenBank baze podataka (pristupni brojevi U90923.1-U90925.1 i KM488213.1 - KM488222.1) kako bi program automatski usporedio alele i dodijelio im nazive poznatih alela ukoliko se podudaraju.

3.8.3. Modul AmpliSAS

Modul AmpliSAS provodi dodatni korak višestrukog sravnjenja (eng. *clustering*) u kojem se prvo odabiru dominantne sekvence (najbrojnije sekvence koje zadovoljavaju odabrane parametre) i zatim se sravnjuju sa svim ostalim sekvencama na temelju čega se prepoznaju artefakti i njihova učestalost se pridodaje sekvencama iz kojih potječu. Ukoliko unutar pojedine grupe postoje sekvence veće učestalosti od one zadane u parametrima, one se mogu svrstati kao subdominantne i odvojiti u zasebnu grupu. Nakon toga se na temelju odabranih parametara uklanjaju artefakti koji nisu nigdje grupirani ili su pogrešno svrstani

kao subdominantne sekvence te se uklanjaju sve sekvence vrlo malih učestalosti (eng. *filtering*) (Sebastian i sur., 2016; Sebastian, 2018).

Za modul AmpliSAS koristila sam sve zadane parametre za tehnologiju Ion Torrent, a zatim sam na pojedinim uzorcima mijenjala parametre grupiranja i filtriranja kako bih vidjela utječu li na promjenu rezultata. Parametar „najmanje učestalosti u odnosu na dominantnu sekvencu“ (eng. *minimum dominant frequency*, MDF) predstavlja graničnu vrijednost za odvajanje sekvence u novu grupu – ukoliko neka sekvenca iz grupe ima istu ili veću učestalost u odnosu na dominantu sekvencu premješta se u zasebnu grupu kao subdominantna sekvenca. Zadana vrijednost mu je 25%, a preporučuje se smanjiti kod jedinki za koje se sumnja da imaju udvostručene lokuse. Parametar „najmanja učestalost amplikona“ (eng. *minimum amplicon frequency*, MAF) uklanja grupe male učestalosti nakon grupiranja i predstavlja granicu za priznavanje pojedine sekvence mogućim pravim alelom, a najmanja preporučena vrijednost mu je 3% (Sebastian i sur., 2016; Sebastian, 2018).

3.8.4. Ostali alati korišteni za analizu sekvenci

Sekvence dobivene u rezultatima modula AmpliCHECK sravnjavala u *online* programima EBMoss Needle i Clustal Omega. Za translaciju nukleotidnih sekvenci koristila sam alat u *online* programskom paketu ExPASy (Swiss Institute of Bioinformatics). Sravnjenja sam preuzela i vizualizirala u programu JalView u kojem sam također izradila filogenetsko stablo metodom sparivanja susjeda (eng. *neighbour joining*, NJ) u kojoj se stablo konstruira na temelju matrice udaljenosti, odnosno broju razlika između nukleotidnih ili proteinskih sekvenci alela (Baxevanis i Oulette, 2001). Kako bih analizirala što više očitavanja, budući da AmpliSAT obrađuje samo prvih 5 000, podijelila sam pročišćene datoteke fastq nekoliko uzoraka u programu Notepad++ na pet dijelova te analizirala svaki od njih zasebno. Ovo se može postići na drugi, jednostavniji način, koristeći računalnu verziju programa AmpliSAT u kojoj je putem naredbe „-s Shuffle/randomize reads/sequences to analyze“ u naredbenom retku (eng. *command line*) moguće primijeniti parametar nasumičnog odabira očitavanja (<https://groups.google.com/g/amplisat/>).

4. REZULTATI

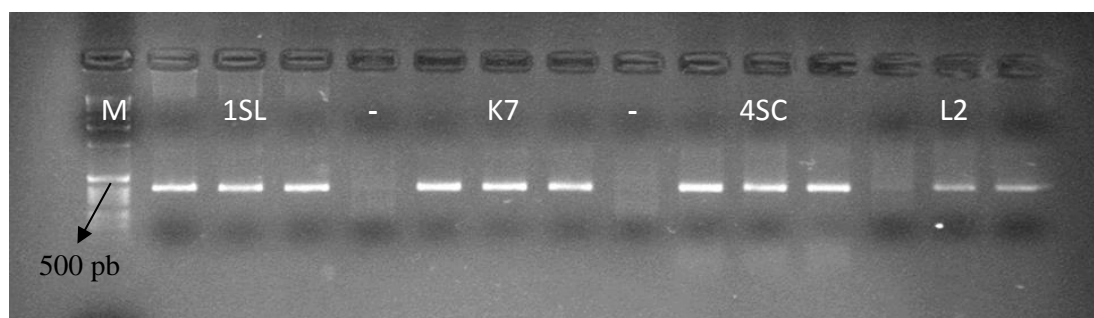
4.1. Koncentracija izolirane DNA

Koncentracije 14 uzoraka izolirane DNA srna bile su u rasponu od 1,5 do 85 ng/ μ L. Uzorak M8/35 je imao koncentraciju ispod razine detekcije te smo ga isključili iz istraživanja.

4.2. Produkti PCR dobiveni nakon umnažanja egzona 2 lokusa DRB

Egzon 2 lokusa DRB uspješno je umnožen kod svih 14 uzoraka; produkt PCR je dobiven u barem dvije od tri replike. Dobivene pruge bile su različitog intenziteta, no za daljnju upotrebu dovoljne su vrlo male količine DNA budući da se knjižnica u konačnici razrjeđuje na manje od 100 pM.

Na Slici 10. prikazani su produkti PCR četiri uzorka. Pruge su očekivane veličine, oko 400 pb, budući da je egzon 2 veličine 270 pb, a adapteri i barkodovi oko 100 pb. Kod zadnjeg uzorka pruge su slabijeg intenziteta zbog manje početne koncentracije DNA, a mogući razlog još slabije koncentracije u prvoj jažici je istjecanje uzorka iz jažice prilikom nanošenja na gel.

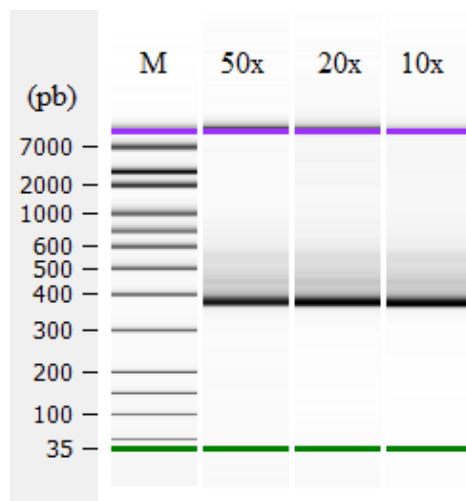


Slika 10. Agarozni gel nakon elektroforeze produkata PCR uzoraka 1SL, K7, 4SC i L2 i molekularnog markera veličina (M). Svaki uzorak ima tri replike i negativnu kontrolu (-), negativne kontrole uzoraka 4SC i L2 nanese su na drugi gel.

4.3 Rezultati provjere kvalitete pripremljene knjižnice na bioanalizatoru

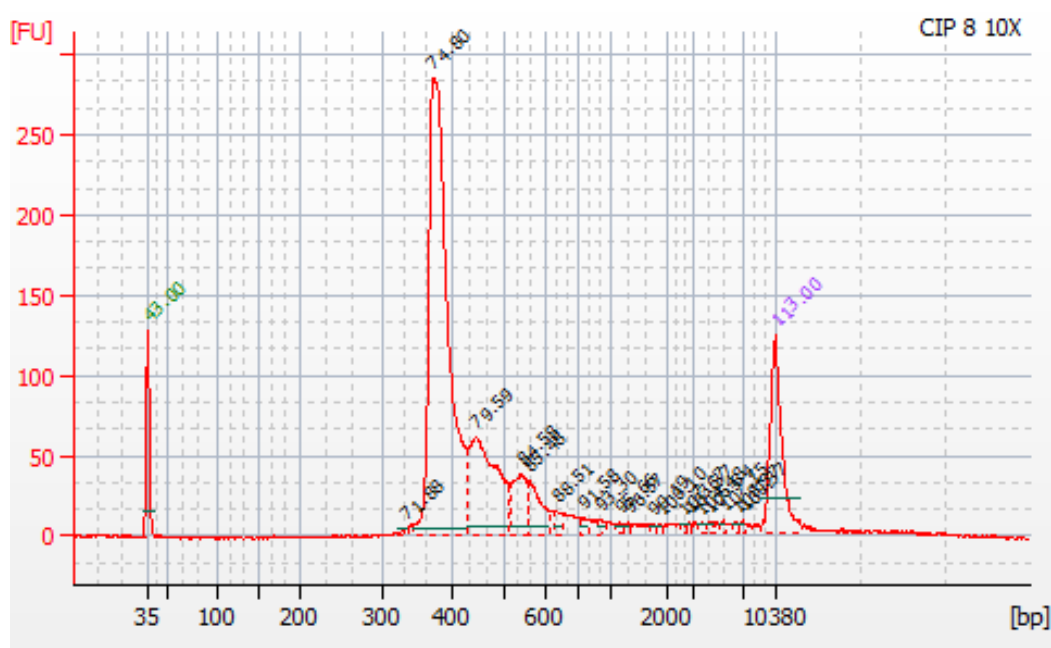
Rezultati bioanalize različitih razrjeđenja pripremljene knjižnice pokazali su koliko je fragmenata DNA u pojedinom razrjeđenju i koje su veličine, čime smo dobili uvid u čistoću i kvalitetu pripremljene knjižnice. Računalni program „2100 Expert Software“ prikazao je rezultate u obliku elektroforetskog gela, elektroferograma i tablice.

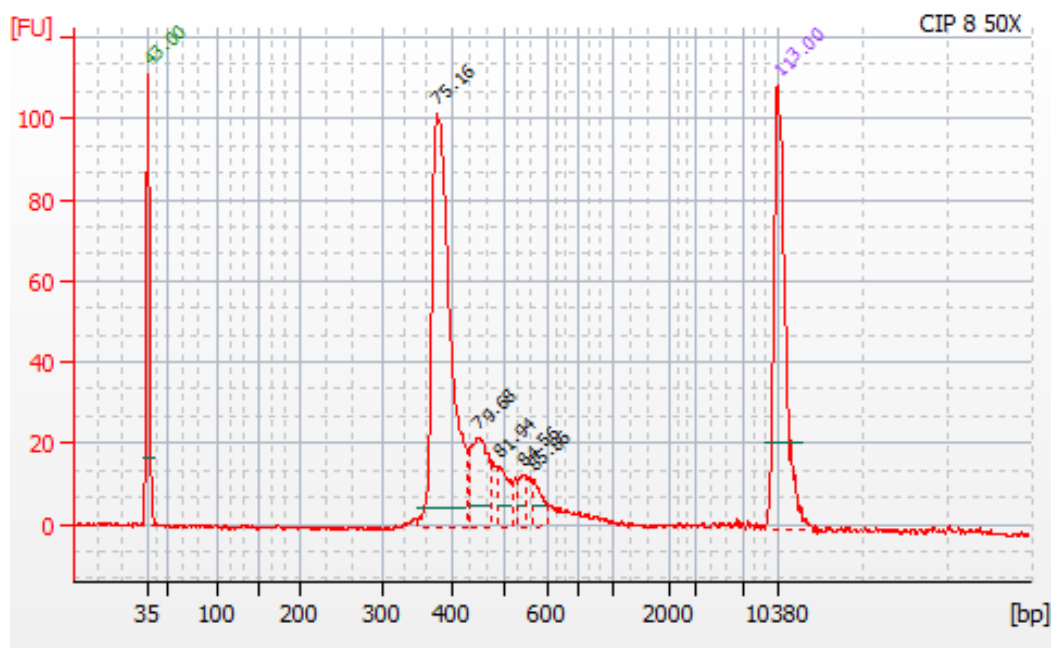
Rezultati koji su prikazani preko slike elektroforetskog gela (Slika 11) nisu vrlo informativni, ali su pogodni za brzu usporedbu svih uzoraka: razrjeđenja knjižnica.



Slika 11. Prikaz elektroforetskog gela. U prvom stupcu prikazan je molekularni marker veličina (M), a u sljedećim razrjeđenja knjižnica. Ljubičaste i zelene pruge predstavljaju standarde.

Najinformativniji prikaz je elektroferogram (Slika 12) iz kojeg možemo vidjeti da u razrjeđenjima prevladavaju fragmenti veličine 360 - 380 pb što su zapravo produkti PCR te da je knjižnica vrlo dobro pročišćena od manjih fragmenata koji potječu od neiskorištenih reagensa, dimera početnica, necjelovitih produkata PCR itd. To je vrlo bitno jer oni mogu biti u kompeticiji sa željenim fragmentima za vezanje na čestice ISP u daljnjim koracima. U 10x razrjeđenju postoji niz fragmenata većih od produkata PCR koji potječu fragmenata DNA kalupa, no oni ne predstavljaju problem jer su u maloj koncentraciji i zbog veličine će se sporije vezati za čestice ISP te će se u konačnici odstraniti. U knjižnici razrijeđenoj 50x takvih fragmenata gotovo da nema zbog velikog razrjeđenja.





Slika 12. Elektroferogrami knjižnica razrijeđenih 10x (iznad) i 50x (ispod). Na x osi su prikazane veličine fragmenata u parovima baza (*base pairs*, bp) od donjeg markera (zeleno), do gornjeg markera (ljubičasto). Na y osi su prikazane jedinice florescencije (*flourescence units*, FU) koje su proporcionalne koncentraciji fragmenata.

Tablični rezultati programa „2100 Expert Software“ vrlo su detaljni i sadrže mnoštvo informacija prema kojima je izrađen elektroferogramski prikaz. U Tablici 4 prikazani su bitniji podaci među kojima je najvažniji podatak molarna koncentracija pojedinog razrjeđenja prema kojoj smo napravili daljnja razrjeđenja knjižnice.

Iako je razrjeđenje 50x „čišće“ što se najbolje vidi na elektroferogramu (Slika 12) za daljnje postupke odabrala sam razrjeđenje 10x kako bi se spriječio mogući gubitak željenih fragmenata prilikom pipetiranja zbog opće male količine fragmenata u razrjeđenju 50x.

Tablica 4. Najvažniji podaci za pojedina razrjeđenja knjižnice iz tabličnih rezultata bioanalize.

| razrjeđenje | prosječna veličina fragmenata (pb) | masena koncentracija (pg/μL) | molarna koncentracija (pmol/L) |
|-------------|------------------------------------|------------------------------|--------------------------------|
| 50x | 443 | 385,21 | 1 379,7 |
| 20x | 443 | 455,6 | 1 634,6 |
| 10x | 445 | 1007,87 | 3 635 |

4.4. Rezultati pripreme za sekvenciranje i sekvenciranja

Priprema za sekvenciranje koja se sastojala od emulzijske reakcije PCR, obogaćivanja čestica ISP i nanošenja uzorka na čip kao i sama reakcija sekvenciranja bile su uspješne (Slika 13). Pokrivenost čipa od 88% je blizu maksimalne vrijednosti, za koju se iz iskustva zna da rijetko kada prelazi 95%. Proces obogaćivanja je u potpunosti uspio i gotovo uopće nema praznih čestica ISP. Od očitavanja pozitivnih čestica ISP 1% čine kontrolne čestice ISP, a ostalo su očitavanja knjižnice. Očitavanja knjižnice prošla su kroz nekoliko razina filtriranja u kojima se uklonilo 10% poliklonalnih očitavanja - očitavanja čestica ISP na kojima je vezano više od jednog fragmenta DNA, i 22% očitavanja niske kvalitete. Od ukupnih 33 363 096 očitavanja, 70% odnosno 23 178 088 očitavanja zadovoljavaju kriterije kvalitete i dostupna su za preuzimanje.



Slika 13. Podaci o uspješnosti reakcije sekvenciranja. Brojevi i postoci odnose se na opis s desne strane, zelenom bojom označen je dio očitavanja koji je prošao pojedine korake filtriranja, sivo su odstranjena očitavanja, a crvena boja predstavlja 1% kontrolnih čestica ISP.

Očitavanja su razvrstana po uzorcima na temelju barkodova. Svi uzorci imaju velik broj očitavanja (195 903 – 376 344) od kojih je više od 90% dobre kvalitete odnosno pripadaju u 70% očitavanja koja zadovoljavaju kriterije kvalitete i monoklonalnosti. Egzon 2 veličine je oko 270 pb, a pojedini uzorci imaju nešto kraću prosječnu duljinu očitavanja što znači da su prisutna manja onečišćenja, ali ona će se u analizi podataka odstraniti.

Tablica 5. Očitavanja pojedinih uzoraka, Q20 je vrijednost koja označava pouzdanost očitavanja od 99% odnosno mogućnost pogrešno očitane baze 1:1000.

| uzorak | broj očitavanja | % očitavanja \geq Q20 | prosječna duljina očitavanja (bp) |
|--------|-----------------|-------------------------|-----------------------------------|
| 1SN | 206 391 | 91,9 | 258 |
| 1SL | 213 718 | 92,1 | 259 |
| 2SL | 376 344 | 91,8 | 257 |
| 3SL | 278 302 | 92,1 | 257 |
| L2 | 310 701 | 92,4 | 271 |
| L5 | 264 817 | 91,9 | 263 |
| L19 | 292 534 | 92,1 | 261 |
| L20 | 211 862 | 90,7 | 270 |
| K7 | 332 978 | 92,1 | 267 |
| K10 | 204 747 | 91,7 | 258 |
| 4SC | 317 923 | 91,5 | 262 |
| 12SC | 360 473 | 92,5 | 266 |
| 13SC | 255 962 | 92,0 | 259 |
| 14SC | 195 903 | 92,1 | 253 |

4.5. Analiza dobivenih očitavanja

U modulu AmpliCLEAN odstranjeno je u prosjeku 77,8% sirovih očitavanja (od 67% do 91%) za svaki uzorak. Točan broj očitavanja za sve uzorke prikazan je u Tablici 1 priloga.

U modulu AmpliCHECK analizirane su pročišćene datoteke i dobila sam informacije o duljini, broju očitavanja i učestalosti pojedine sekvence u uzorku. U rezultatima su prikazane samo sekvence sa učestalošću više od 1%. U prosjeku su za jedan uzorak pronađene četiri različite sekvence, najviše ih je pronađeno šest, a za nekoliko njih samo dvije dominantne sekvence imaju učestalost $>1\%$. Prema opisanom mehanizmu nastanka najveći dio artefakata čine delecijски ili insercijski artefakti koji se od alela razlikuju za jednu bazu više ili manje, a najmanje supstitucijski artefakti. Najveća učestalost artefakta je 4,94%, a najmanja učestalost alela 11,04%. Aleli su dugački 267 ili 270 pb.

Kao primjer prikazala sam rezultate modula AmpliCHECK za uzorak 14SC (Tablica 6). Pronađeno je četiri različitih sekvenci učestalosti veće od 1%. Sekvence s najvećom učestalošću i brojem očitavanja podudaraju se s alelima iz baze podataka. Zadnje dvije sekvence su prisutne u malom postotku i vjerojatno su artefakti dominantnih sekvenci. Sekvenca MHC2-0005 mogla je nastati ili kao kimera prve dvije sekvence ili jednom supstitucijskom mutacijom sekvence Caca-DRB*0102.

Zadnja sekvenca se od alela CacaDRB1*0301 razlikuje po jednoj homopolimernoj inserciji. Sekvence s učestalošću manje od 1% nisu prikazane i čine čak 47,02% ukupnih očitavanja ovog uzorka.

Tablica 6. Najvažniji podaci iz rezultata AmplicHECK modula za uzorak 14SC. Zelenom bojom označene su dominantne sekvence, a crvenom bojom sekvence koje su najvjerojatnije artefakti i uz njih je opisan mogući mehanizam nastanka: CH - kimerne sekvence, X – substitucije, D - delecije, I – insercije, H – homopolimerne indel.

| sekvenca | duljina (pb) | broj očitavanja | učestalost (%) | mogući mehanizmi nastanka artefakta |
|---------------|--------------|-----------------|----------------|---|
| Caca-DRB*0102 | 270 | 1593 | 31,86 | |
| CacaDRB1*0301 | 267 | 938 | 18,76 | |
| MHC2-0005 | 270 | 62 | 1,24 | CH: CacaDRB1*0301+ Caca-DRB*0102, X1: Caca-DRB*0102 |
| CacaDRB1*0301 | 268 | 56 | 1,12 | I1H1: CacaDRB1*0301 |

Svi artefakti prikazani u rezultatima modula AmplicHECK s opisanim mehanizmom mutacije u homopolimernoj regiji sadrže inserciju (I1H1) i pronašla sam ih kod šest uzoraka. Višestruko sravnjenje sekvenci tih artefakata i alela iz kojih potječu prikazano je na Slici 14. Sve homopolimerne insercije nalaze se u istoj regiji: jedinoj regiji od pet uzastopnih nukleotida koja se nalazi na području 187-191 pb. U nekoliko područja s četiri uzastopna nukleotida nisam pronašla polimorfna mjesta.

| | 170 | 180 | 190 | 200 | 210 | 220 | 230 | 240 |
|-------------------------|----------------------|-----------------------|--------------|-------------|-----------|-----|-----|-----|
| 1SL_MHC-002AACAGCCAGA | AGGAGTTCTGGAGCGGG- | GGCGGGCCGCGGTGGACACGT | ACTGCAGACACA | AACTACGGGGT | CGGTGAGAG | | | |
| L2_MHC-003AACAGCCAGA | AGGAGTTCTGGAGCGGG- | GGCGGGCCGCGGTGGACACGT | ACTGCAGACACA | AACTACGGGGT | TATTGAGAG | | | |
| 1SL_I1H1/1-27AACAGCCAGA | AGGAGTTCTGGAGCGGG- | GGCGGGCCGCGGTGGACACGT | ACTGCAGACACA | AACTACGGGGT | CGGTGAGAG | | | |
| L2_I1H1/1-271AACAGCCAGA | AGGAGTTCTGGAGCGGG- | GGCGGGCCGCGGTGGACACGT | ACTGCAGACACA | AACTACGGGGT | TATTGAGAG | | | |
| 14SC_I1H1/1-1AACAGCCAGG | - - AGATCCTGGAGCGGG- | GGCGGGCCGCGGTGGACACGT | ACTGCAGACACA | AACTACGGGGT | TATTGAGAG | | | |
| L19_I1H1/1-26AACAGCCAGG | - - AGATCCTGGAGCGGG- | GGCGGGCCGCGGTGGACACGT | ACTGCAGACACA | AACTACGGGGT | TATTGAGAG | | | |
| L20_I1H1/1-26AACAGCCAGG | - - AGATCCTGGAGCGGG- | GGCGGGCCGCGGTGGACACGT | ACTGCAGACACA | AACTACGGGGT | TATTGAGAG | | | |
| 14SC_U90925AACAGCCAGG | - - AGATCCTGGAGCGGG- | GGCGGGCCGCGGTGGACACGT | ACTGCAGACACA | AACTACGGGGT | TATTGAGAG | | | |
| L19_U90925.1AACAGCCAGG | - - AGATCCTGGAGCGGG- | GGCGGGCCGCGGTGGACACGT | ACTGCAGACACA | AACTACGGGGT | TATTGAGAG | | | |
| L20_KM48821AACAGCCAGG | - - AGATCCTGGAGCGGG- | GGCGGGCCGCGGTGGACACGT | ACTGCAGACACA | AACTACGGGGT | TATTGAGAG | | | |

Slika 14. Višestruko sravnjenje artefakata za koje je opisani mehanizam nastanka homopolimerna insercija (I1H1) i alela iz kojih proizlaze. Prikazan je samo dio sravnjenja koji sadži indel mutacije i one su istaknute tamno sivom bojom, dok su ostale pozicije obojane svijetlo sivom bojom. Na osi x prikazane su nukleotidne pozicije, a na y osi naziv uzorka.

Rezultati modula AmpliSAS dobiveni uz zadane parametre za tehnologiju Ion Torrent prikazani su u Tablici 7 i 8. U 14 analiziranih jedinki nađeno je osam različitih alela. Od ukupno 13 alela lokusa DRB srna koji se trenutno nalaze u bazi podataka GenBank ovdje ih je pronađeno pet (CacaDRB1*0301, Caca-DRB*0102, *0302*, *0304 i *0401) Tri alela radnih naziva MHC-0001, MHC-0002 i MHC-0003 su nova, od kojih se aleli MHC-0001 i MHC-0002 poklapaju s alelima nađenim u istraživanju slovenskih populacija srna (Bužan, neobjavljeni rezultati), a alel MHC-0003 potpuno je nov. Najučestaliji alel je Caca-DRB*0102 koji se pojavljuje kod sedam jedinki, a najmanje su zastupljeni novi aleli MHC-0002 i MHC-0003 (Tablica 7).

Tablica 7. Broj očitavanja pojedinih alela, ukupnih alela po uzorku (dubina alela, eng. *allele depth*) i ukupnih očitavanja (dubina amplikona, eng. *amplicon depth*). Udio u ukupnom uzorku predstavlja broj jedinki kod kojih je nađen određeni alel.

| | 1SN | 1SL | 2SL | 3SL | L2 | L5 | L19 | L20 | K7 | K10 | 4SC | 12SC | 13SC | 14SC | Udio |
|------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| CacaDRB1*0301 | | | | | | 1015 | 1148 | 1337 | | | | 1994 | 1496 | 1737 | 6 |
| Caca-DRB*0302 | 2092 | | | | | | | 3001 | | 2404 | 2200 | | | | 4 |
| Caca-DRB*0102 | 2092 | 1701 | 1687 | 1687 | | | | | | 1833 | 2023 | | | | 7 |
| Caca-DRB*0304 | | | | | | | | | 2717 | | | 3379 | | 2524 | 3 |
| Caca-DRB*0401 | | | | | | | | | | | | | 2748 | | 2 |
| MHC-0001 | | | 2374 | | 2069 | | | | 1412 | | | | | | 2 |
| MHC-0002 | | 2567 | | 2392 | | 3192 | 3010 | | | | | | | | 1 |
| MHC-0003 | | | | | 2367 | | | | | | | | | | 1 |
| dubina alela* | 4124 | 4434 | 4061 | 4079 | 4436 | 4207 | 4158 | 4338 | 4129 | 4237 | 4229 | 4373 | 4244 | 4261 | |
| dubina amplikona | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | |

Iako su svi uzorci heterozigoti postoje odstupanja od očekivanog omjera 1:1. Najveće odstupanje pokazuje uzorak 12SC s učestalosti alela 67,58 % i 19,88 %. Uzorci L5, L19 i L20 također pokazuju omjer alela ~1:3 i kod svih je manje zastupljen alel CacaDRB1*0301 (Tablica 8) što bi moglo značiti da se slabije umnaža korištenim početnicama.

Tablica 8. Najvažniji rezultati modula AmpliSAS dobiveni zadanim parametrima za tehnologiju sekvenciranja Ion Torrent.

| uzorak | naziv alela | GenBank pristupni broj | duljina (pb) | broj očitavanja | učestalost alela u uzorku (%) |
|--------|---------------|---------------------------|-----------------|-----------------|-------------------------------------|
| 1SN | Caca-DRB*0302 | KM488213.1 | 267 | 2 092 | 41,84 |
| | Caca-DRB*0102 | KM488214.1 | 270 | 2 032 | 40,64 |
| 1SL | MHC-0002 | / | 270 | 2 567 | 51,34 |
| | Caca-DRB*0102 | KM488214.1 | 270 | 1 701 | 34,02 |
| 2SL | MHC-0001 | / | 267 | 2 374 | 47,48 |
| | Caca-DRB*0102 | KM488214.1 | 270 | 1 687 | 33,74 |
| 3SL | MHC-0002 | / | 270 | 2 392 | 47,84 |
| | Caca-DRB*0102 | KM488214.1 | 270 | 1 687 | 33,74 |
| L2 | MHC-0003 | / | 270 | 2 367 | 47,34 |
| | MHC-0001 | / | 267 | 2 069 | 41,38 |
| L5 | Caca-DRB*0401 | KM488218.1 | 270 | 3 192 | 63,84 |
| | CacaDRB1*0301 | U90925.1 | 267 | 1 015 | 20,3 |
| L19 | Caca-DRB*0401 | KM488218.1 | 270 | 3 010 | 60,2 |
| | CacaDRB1*0301 | U90925.1 | 267 | 1 148 | 22,96 |
| L20 | Caca-DRB*0302 | KM488213.1 | 267 | 3 001 | 60,02 |
| | CacaDRB1*0301 | U90925.1 | 267 | 1 337 | 26,74 |
| K7 | Caca-DRB*0304 | KM488215.1 | 270 | 2 717 | 54,34 |
| | MHC2-0001 | / | 267 | 1 412 | 28,24 |
| K10 | Caca-DRB*0302 | KM488213.1 | 267 | 2 404 | 48,08 |
| | Caca-DRB*0102 | KM488214.1 | 270 | 1 833 | 36,66 |
| 4SC | Caca-DRB*0302 | KM488213.1 | 267 | 2 200 | 44 |
| | Caca-DRB*0102 | KM488214.1 | 270 | 2 023 | 40,58 |
| 12SC | Caca-DRB*0304 | KM488215.1 | 270 | 3 379 | 67,58 |
| | CacaDRB1*0301 | U90925.1 | 267 | 1 994 | 19,88 |
| 13SC | Caca-DRB*0401 | KM488218.1 | 270 | 2 748 | 54,96 |
| | CacaDRB1*0301 | U90925.1 | 267 | 1 496 | 29,92 |
| 14SC | Caca-DRB*0102 | KM488214.1 | 270 | 2 524 | 50,48 |
| | CacaDRB1*0301 | U90925.1 | 267 | 1 737 | 34,74 |

Promjena parametara ni u kojem slučaju nije utjecala na konačne rezultate: određivanje alela jedinki. Neznatne promjene učestalosti alela dobila sam primjenom parametara prikazanih u Tablici 9. Smanjenjem MDF na 1 i manje neke sekvence su se odvojile u zasebne grupe i učestalost dominantnih sekvenci se smanjila, što bi značilo da unutar grupe nema sekvenci čija je učestalost u odnosu na dominantne sekvence veća od 1%. Isto tako sve su sekvence unutar grupe vrlo slične dominantnoj sekvenci te se učestalosti smanjuju kod granične vrijednosti sličnosti od oko 99%.

Tablica 9. Učestalosti alela uzoraka 3SL, 12SC i L2 kod promjene parametara najmanje učestalosti u odnosu na dominantnu sekvencu (eng. *minimum dominant frequency*, MDF) i postotka sličnosti (eng. *clustering identity threshold*, CIT). Crvenom bojom su označene promijenjene učestalosti alela.

| uzorak | alel | parametar | | | | | |
|--------|---------------|-----------|-------|--------|---------|-------|-------|
| | | MDF (%) | | | CIT (%) | | |
| | | 0.5 | 1 | 5 - 25 | 98 | 99 | 99.5 |
| 3SL | MHC-0002 | 44,7 | 47,84 | 47,84 | 47,84 | 47,72 | 46,7 |
| | Caca-DRB*0102 | 32,44 | 33,74 | 33,74 | 33,74 | 33,6 | 33,16 |
| 12SC | Caca-DRB*0304 | 65,74 | 67,58 | 67,58 | 67,58 | 67,42 | 65,74 |
| | CacaDRB1*0301 | 17,62 | 19,24 | 19,88 | 19,88 | 19,86 | 19,46 |
| L2 | MHC-0003 | 44,56 | 46,32 | 47,34 | 47,34 | 47,06 | 46,14 |
| | MHC-0001 | 39,02 | 40,68 | 41,38 | 41,38 | 41,28 | 39,16 |

Veće promjene učestalosti alela dobila sam analizom podijeljenih datoteka fastq, no ni tada nije dobivena promjena konačnih rezultata – alel koji je najbrojniji u prvom dijelu najbrojniji je i u svakom drugom analiziranom dijelu datoteke. U Tablici 10 su prikazani rezultati analize za uzorke 3SL i 12SC.

Tablica 10. Učestalosti alela uzoraka 3SL i 12SC u različitim dijelovima datoteke fastq. Crvenom bojom je označena najveća razlika u omjeru alela, a zelenom najmanja.

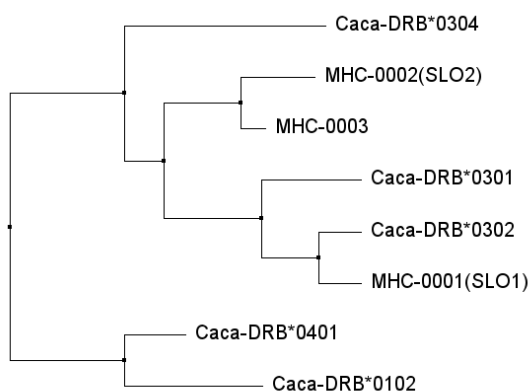
| uzorak | dio | alel | učestalost alela |
|--------|-----|---------------|------------------|
| 3SL | 1. | MHC-0002 | 47,84 |
| | | Caca-DRB*0102 | 33,74 |
| | 2. | MHC-0002 | 49,76 |
| | | Caca-DRB*0102 | 30 |
| | 3. | MHC-0002 | 51,12 |
| | | Caca-DRB*0102 | 30,9 |
| | 4. | MHC-0002 | 48,98 |
| | | Caca-DRB*0102 | 31,1 |
| | 5. | MHC-0002 | 46,04 |
| | | Caca-DRB*0102 | 34,98 |

| | | | |
|------|----|---------------|-------|
| 12SC | 1. | Caca-DRB*0304 | 67,58 |
| | | CacaDRB1*0301 | 19,88 |
| | 2. | Caca-DRB*0304 | 74,8 |
| | | CacaDRB1*0301 | 12,32 |
| | 3. | Caca-DRB*0304 | 70 |
| | | CacaDRB1*0301 | 18,22 |
| | 4. | Caca-DRB*0304 | 66,6 |
| | | CacaDRB1*0301 | 20,04 |
| | 5. | Caca-DRB*0304 | 71,44 |
| | | CacaDRB1*0301 | 16,5 |

Višestrukim sravnjenjem osam pronađeno je 26 polimorfniha mjesta od kojih 23 sadrže supstitucijske mutacije, a tri deleciju (Slika 15). Evolucijski odnosi sekvenci koji prikazani su na Slici 16. Sekvence MHC-0002 i MHC-0003 razlikuju se u samo jednom nukleotidu i međusobno su najslbližnije. Tri sekvence duge su 267 pb i možemo ih grupirati zajedno po sličnosti; toj skupini pripada i novi alel MHC-0001 koji se od Caca-DRB*0302 također razlikuje samo po jednoj bazi.

| | | | | | | |
|----------------|---|-----------------------------------|---|----------------------|-----|-----|
| | 10 | 20 | 30 | 40 | 50 | 60 |
| Caca-DRB*0401/ | GATGTATACTAAGAAAGAGTGT | TATTTCTCCAAT | GGGACGGAGCGGGT | GCGGTTCTTGACAGATACTT | | |
| Caca-DRB*0102/ | GATGTATACTACGGGCGAGTGT | CATTTCTCCAAC | GGGACGGAGCGGGT | GCGGTTCTTGACAGATACTT | | |
| Caca-DRB*0304/ | GATGTATACTACGAGCGAGTGT | CATTTCTCCAAC | GGGACGGAGCGGGT | GCGGTTCTTGACAGATACTT | | |
| MHC-0002(SLO2) | GATGTATACTAAGAAAGAGTGT | TATTTCTCCAAT | GGGACGGAGCGGGT | GCGGTTCTTGACAGATACTT | | |
| MHC-0003/1-270 | GATGTATACTAAGAAAGAGTGT | TATTTCTCCAAT | GGGACGGAGCGGGT | GCGGTTCTTGACAGATACTT | | |
| Caca-DRB*0301/ | GATGTATACTACGAGCGAGTGT | CATTTCTCCAAC | GGGACGGAGCGGGT | GCGGTTCTTGACAGATACTT | | |
| Caca-DRB*0302/ | GATGTATACTAAGAAAGAGTGT | TATTTCTCCAAC | GGGACGGAGCGGGT | GCGGTTCTTGACAGATACTT | | |
| MHC-0001(SLO1) | GATGTATACTAAGAAAGAGTGT | TATTTCTCCAAT | GGGACGGAGCGGGT | GCGGTTCTTGACAGATACTT | | |
| | 70 | 80 | 90 | 100 | 110 | 120 |
| Caca-DRB*0401/ | TTCTATAACGGAGAAGAGT | TCGTGCGCTTCGACAGCGACT | GGGGCGAGTACCGGGCGGTGACCGAGCTG | | | |
| Caca-DRB*0102/ | TTCTATAACGGAGAAGAGT | TCGTGCGCTTCGACAGCGACT | GGGGCGAGTACCGGGCGGTGACCGAGCTG | | | |
| Caca-DRB*0304/ | TTCTATAACGGAGAAGAGT | TCGTGAGCTTCGACAGCGACT | GGGGCGAGTACCGGGCGGTGACCGAGCTG | | | |
| MHC-0002(SLO2) | TTCTATAACGGAGAAGAGT | TCGTGCGCTTCGACAGCGACT | GGGGCGAGTACCGGGCGGTGACCGAGCTG | | | |
| MHC-0003/1-270 | TTCTATAACGGAGAAGAGT | TCGTGCGCTTCGACAGCGACT | GGGGCGAGTACCGGGCGGTGACCGAGCTG | | | |
| Caca-DRB*0301/ | TTCTATAACGGAGAAGAGT | TCGTGCGCTTCGACAGCGACT | GGGGCGAGTACCGGGCGGTGACCGAGCTG | | | |
| Caca-DRB*0302/ | TTCTATAACGGAGAAGAGT | TCGTGCGCTTCGACAGCGACT | GGGGCGAGTACCGGGCGGTGACCGAGCTG | | | |
| MHC-0001(SLO1) | TTCTATAACGGAGAAGAGT | TCGTGCGCTTCGACAGCGACT | GGGGCGAGTACCGGGCGGTGACCGAGCTG | | | |
| | 140 | 150 | 160 | 170 | 180 | 190 |
| Caca-DRB*0401/ | GGCGGCCCGGTGGCCGAGGGCTGGAACAGCCAGAAGGAGT | TCCTGGAGCAGAGGCGGGCCGAGGTGGACA | | | | |
| Caca-DRB*0102/ | GGCGGCCCGGTGGCCGAGGGCTGGAACAGCCAGAAGGAGT | TCCTGGAGCAGAGGCGGGCCGAGGTGGACA | | | | |
| Caca-DRB*0304/ | GGCGGCCCGTCCGCCAAGTACTGGAACAGCCAGAAGGAGAT | CCTGGAGCAGAGGCGGGCCGCGGTGGACA | | | | |
| MHC-0002(SLO2) | GGCGGCCCGGACGCCAAGTACTGGAACAGCCAGAAGGAGT | TCCTGGAGCGGGGGCGGGCCGCGGTGGACA | | | | |
| MHC-0003/1-270 | GGCGGCCCGGACGCCAAGTACTGGAACAGCCAGAAGGAGT | TCCTGGAGCGGGGGCGGGCCGCGGTGGACA | | | | |
| Caca-DRB*0301/ | GGCGGCCCGACGCCAAGTACTGGAACAGCCAGG | AGATCCTGGAGCGGGGGCGGGCCGCGGTGGACA | | | | |
| Caca-DRB*0302/ | GGCGGCCCGACGCCAAGTACTGGAACAGCCAGG | AGATCCTGGAGCGGGGGCGGGCCGCGGTGGACA | | | | |
| MHC-0001(SLO1) | GGCGGCCCGACGCCAAGTACTGGAACAGCCAGG | AGATCCTGGAGCGGGGGCGGGCCGCGGTGGACA | | | | |
| | 210 | 220 | 230 | 240 | 250 | 260 |
| Caca-DRB*0401/ | GGACACGTACTGCAGACACAAC | TACGGGGT | CGGTGAGAGTTTCACTGTGCAGCGGCGAGGTGACGCGAA | | | |
| Caca-DRB*0102/ | GGACACGTACTGCAGACACAAC | TACGGGGT | CGGTGAGAGTTTCACTGTGCAGCGGCGAGGTGACGCGAA | | | |
| Caca-DRB*0304/ | GGACACGTACTGCAGACACAAC | TACGGGGT | TATTGAGAGTTTCACTGTGCAGCGGCGAGGTGACGCGAA | | | |
| MHC-0002(SLO2) | GGACACGTACTGCAGACACAAC | TACGGGGT | CGGTGAGAGTTTCACTGTGCAGCGGCGAGGTGACGCGAA | | | |
| MHC-0003/1-270 | GGACACGTACTGCAGACACAAC | TACGGGGT | TATTGAGAGTTTCACTGTGCAGCGGCGAGGTGACGCGAA | | | |
| Caca-DRB*0301/ | GGACACGTACTGCAGACACAAC | TACGGGGT | TATTGAGAGTTTCACTGTGCAGCGGCGAGGTGACGCGAA | | | |
| Caca-DRB*0302/ | GGACACGTACTGCAGACACAAC | TACGGGGT | TATTGAGAGTTTCACTGTGCAGCGGCGAGGTGACGCGAA | | | |
| MHC-0001(SLO1) | GGACACGTACTGCAGACACAAC | TACGGGGT | TATTGAGAGTTTCACTGTGCAGCGGCGAGGTGACGCGAA | | | |

Slika 15. Višestruko sravnjenje nukleotidnih sekvenci obojano prema postotku sličnosti (eng. *percentage identity*) u programu JalView. Na osi x prikazane su nukleotidne pozicije, a na y osi naziv alela. Nijanse označavaju postotak identičnih nukleotida na određenom položaju: tamno plava - svi nukleotidi su identični, nijanse svijetlo plave boje 50-100% nukleotida je identično, bijela boja < 50% nukleotida je identično.



Slika 16. Filogenetsko stablo temeljeno na višestrukome sravnjenju iz prethodne slike, konstruirano metodom *neighbour joining* u programu JalView.

Višestrukim sravnjenjem aminokiselinskih sekvenci nađeno je 14 varijabilnih mjesta (Slika 17). Od ukupno 23 supstitucije u nukleotidnim sekvencama osam ih je sinonimno, a 13 nesinonimno: rezultiraju promjenom aminokiseline. Delecija tri uzastopna nukleotida, odnosno kodona na poziciji 65 rezultira delecijom jedne aminokiseline te ne remeti okvir čitanja. Od sekvenci koje se međusobno razlikuju u samo jednoj bazi, MHC-0002 i MHC-0003 se razlikuju također u jednoj aminokiselini, a sekvence MHC-0001 i Caca-DRB*0302 daju identične polipeptidne lance, stoga osam različitih alela nađenih u ovom istraživanju čini sedam različitih aminokiselinskih sekvenci.



Slika 17. Višestruko sravnjenje aminokiselinskih sekvenci obojano prema postotku sličnosti u programu JalView. Na osi x prikazane su aminokiselinske pozicije, a na y osi naziv alela. Nijanse označavaju postotak identičnih aminokiselina na određenom položaju: tamno plava - sve aminokiseline su identične, nijanse svijetlo plave boje 50-100% aminokiseline je identično, bijela boja < 50% aminokiseline je identično.

5. RASPRAVA

Korake u ovom istraživanju, i općenito u genotipizaciji lokusa korištenjem metoda sekvenciranja nove generacije, NGS, mogla bih podijeliti u dva dijela: eksperimentalni dio pripreme DNA fragmenata za sekvenciranje i bioinformatički dio analize dobivenih očitavanja. Priprema za sekvenciranje sastoji se od mnogo koraka i uspješnost svakog od njih utječe na konačne rezultate sekvenciranja, no analiza podataka ne mora nužno ovisiti o kvaliteti pripremljene knjižnice. U ovom istraživanju je broj barkodova koje sam imala na raspolaganju bio manji od maksimalnog broja uzoraka koji se mogu sekvencirati na Ion™ 530 čipu (nekoliko tisuća, ovisno o pokrivenosti i duljini fragmenta, <https://www.thermofisher.com/>) te sam zbog malog broja uzoraka (95) i velike pokrivenosti čipa (88%) dobila izrazito velik broj očitavanja za svaki uzorak (Tablica 5). Kada bi znatan postotak očitavanja bio isfiltriran zbog poliklonalnosti, loše kvalitete ili velike količine onečišćenja očitavanja bi i dalje bilo dovoljno za pouzdanu genotipizaciju. Ako se na čip nanosi maksimalan broj uzoraka, što je najisplativije, tada priprema za sekvenciranje treba biti što preciznija i potrebno je raditi dodatne kontrole kvalitete kako zbog manjka očitavanja ne bi morali ponavljati cijeli postupak sekvenciranja već samo onaj korak koji nije uspio.

Priprema knjižnice u ovom istraživanju bila je jednostavna; za PCR nije bila potrebna DNA visoke koncentracije ni kvalitete budući da je egzon 2 lokusa DRB manji od 300 pb i za daljnje korake su bile dovoljne vrlo male količine produkta PCR jer se knjižnica fragmenata za sekvenciranje u konačnici razrjeđuje na manje od 100 pM. Produkti PCR su dodani u zajedničku knjižnicu tako da im koncentracije budu otprilike jednake; nije bilo potrebno precizno podešavanje koncentracije jer smo očekivali velik broj očitavanja za svaki uzorak. U slučaju sekvenciranja velikog broja uzoraka bilo bi bolje točno izračunati volumen svakog produkta PCR kojeg je potrebno dodati u knjižnicu kako bi koncentracije svakog produkta PCR bile jednake i time svaki uzorak imao podjednak, i dovoljan, broj očitavanja. Produkti PCR su prethodno dobro pročišćeni od manjih fragmenata koji bi mogli zauzimati mjesto na česticama ISP, što je potvrđeno u bioanalizi. Svako od tri razrjeđenja analiziranih na bioanalizatoru (10x, 20x i 50x) moglo je poslužiti za daljnje korake prema kriteriju čistoće (Slika 11), no odabrali smo najmanje razrjeđenje kako bi izbjegli mogućnost gubitka fragmenata zbog njihove nasumične raspodjele u trenutku pipetiranja kod velikog razrjeđenja. Za daljnje korake emulzijske reakcije PCR, obogaćivanja sfera i nanošenja na čip najvažnije je detaljno proučiti upute prije izvođenja i napraviti plan rada kako se ne bi

nepotrebno zaustavljali na vremenski ograničenim koracima i time umanjili mogućnost njihovog uspješnog provođenja.

Reakcija sekvenciranja je bila uspješna (Slika 13): u emulzijskoj reakciji PCR nastao je dio poliklonalnih čestica ISP (10%), što se može umanjiti pažljivim razrjeđenjem knjižnice, no ne može se u potpunosti izbjeći jer čak i kod velikih razrjeđenja knjižnice dolazi do ulaska više od jednog fragmenta u pojedinu kapljicu emulzije. Obogaćivanje čestica ISP bilo je u potpunosti uspješno i gotovo da nije bilo praznih sfera, što pridonosi, ne samo broju, već i kvaliteti očitavanja. Uspješnost emulzijske reakcije PCR i obogaćivanja sfera može se provjeriti po tome imamo li u završnoj mikroepruveti dovoljan volumen reakcijske otopine, ali mogu se napraviti i dodatni koraci provjere kvalitete kako bi točno znali postotak uspješnosti pojedine reakcije. Ovdje ti koraci nisu rađeni, ali u slučaju sekvenciranja velikog broja uzoraka bilo bi dobro napraviti dodatne provjere kako ne bi morali ponavljati cijeli postupak sekvenciranja zbog manjka očitavanja. Najviše očitavanja u ovom slučaju odstranjeno je zbog loše kvalitete (22%), što može ovisiti o samom sustavu detekcije i reagensima, budući da su svi ostali koraci bili uspješni. Odstranjeno je ukupno 30% očitavanja, no očitavanja po uzorku i dalje ima puno više nego što se u računalnoj analizi korištenjem programskog paketa AmpliSAT može obraditi. AmpliSAT može obraditi 5 000 očitavanja po uzorku (Sebastian i sur., 2016), što je u ovom istraživanju činilo od 1,3% do 2,5% ukupnih očitavanja po uzorku, s obzirom da je broj očitavanja po uzorku iznosio između 195 903 i 360 473 (Tablica 5). Čak i nakon dodatnog pročišćavanja datoteka u modulu AmpliCLEAN u kojem je za neke uzorke odstranjeno više od 80% očitavanja (Prilog 1), ostalo ih je i više nego dovoljno i pokrivenost svakog uzorka je bila dovoljna za pouzdano određivanje baza.

Analiza očitavanja, dakle, može i ne mora ovisiti o kvaliteti pripreme knjižnice. Najveće izazove u analizi očitavanja i genotipizaciji predstavljaju artefakti koji neizbježno nastaju u reakciji PCR. Artefakti nastaju i neispravnim očitanjem baze u procesu sekvenciranja i takvi artefakti zajedno čine veliki postotak svih očitavanja, ali su pojedinačne sekvence tih artefakata obično malih učestalosti i lako se uklanjaju (Paarek i sur., 2011; Babik, 2010). Kao što je prikazano u rezultatima modula AmpliCHECK (Tablica 6) artefakti mogu činiti 50% ukupnih očitavanja i mogu biti većih učestalosti, ako nastanu u ranijim ciklusima reakcije PCR. To predstavlja veliki problem u analizi vrsta koje imaju udvostručene lokuse koje istražujemo, jer su i pravi aleli ponekad male učestalosti pa se razlikovanje ne može temeljiti na pretpostavci da su artefakti manje zastupljeni od pravih alela, kao što je to u slučaju vrsta koje nemaju udvostručene lokuse. Budući da srne imaju samo jedan lokus DRB što je potvrđeno u ranijim

istraživanjima (Mikko i sur., 1999; Van den Bussche i sur., 1999; Quemere i sur., 2015) očekivali smo najviše dva alela po jedinci tako da je razlikovanje pravih alela od artefakata u ovom slučaju poprilično jednostavno.

Već iz rezultata dobivenih analizom u modulu AmpliCHECK mogla sam zaključiti da sve jedinke imaju dva alela i da su njihove učestalosti znatno veće od učestalosti artefakata. Prema najmanjoj učestalosti alela i najvećoj učestalosti artefakta mogla sam u parametrima filtriranja za modul AmpliSAS postaviti vrijednost „najmanja učestalost amplikona“ (eng. *minimum amplicon frequency*, MAF) između 5 i 10 kako bi se uklonili svi artefakti. Budući da svi uzorci imaju dva alela parametar najmanje učestalosti u odnosu na dominantnu sekvencu (eng. *minimum dominant frequency*, MDF) ostavila sam na 25%. Isto tako vidjela sam i da je duljina pravih alela različita – neki su dugi 267, a neki 270 pb te za parametar „obavezna točna duljina očitavanja“ (eng. *exact length required*) ne smijem staviti „da“, iako je za tehnologiju Ion Torrent to preporučeno zbog sklonosti krivog očitavanja homopolimernih regija (Sebastian i sur., 2016).

Zbog iste se smatra da je tehnologija Ion Torrent manje pouzdana od tehnologije Illumina čije su greške u potpunosti nasumične (Pareek i sur., 2011). Višestrukim sravnjem sekvenci artefakata iz rezultata modula AmpliCHECK za koje je opisan mehanizam nastanka homopolimerna insercija i alela iz kojih proizlaze primijetila sam da se sve homopolimerne insercije nalaze na istom mjestu: u području 187 -190 pb s pet uzastopnih gvanina (Slika 14) koja je ujedno i jedina regija s pet uzastopnih nukleotida u alelima, a nalazi se kod pet od osam alela nađenih u ovom istraživanju (Slika 15). Ne možemo znati jesu li svi ti artefakti posljedica krivog očitavanja tehnologije Ion Torrent ili su nastali u reakciji PCR, ali mogli bih zaključiti da pet uzastopnih gvanina predstavlja prag od kojeg greške postaju nenasumične i učestale, jer kod nekoliko regija od četiri uzastopna nukleotida nisam pronašla polimorfna mjesta. Ipak, učestalosti takvih artefakata su u ovom slučaju manje od 3%, a artefakti s najvećim učestalostima ne pripadaju toj skupini tako da za genotipizaciju vrste bez dupliciranih lokusa i bez dužih homopolimernih regija artefakti nastali krivim očitanjem baza sustava Ion Torrent ne bi trebali predstavljati problem.

U modulu AmpliSAS dobila sam iste alele kao i u modulu AmpliCHECK, ali različitim učestalosti zbog toga što su učestalosti artefakta pridodane učestalostima alela iz kojih potječu. Omjeri alela u rezultatima modula AmpliSAS i AmpliCHECK za neke uzorke znatnije variraju što može značiti da određeni alel ima više artefakata ili su oni nastali ranije u

reakciji PCR. Rezultati koje nisam očekivala su veća odstupanja od predviđenog omjera učestalosti alela za heterozigote, a to je 1:1 (Tablica 8). Samo uzorci 1SN i 4SC imaju omjer alela otprilike 1:1. Uzorci L5, L19, L20 i 12SC imaju omjere alela otprilike 1:3. U radu Stipoljev i sur. (2020) u kojem je analiziran egzon 2 lokusa DRB kod divokoze u programskom paketu AmpliSAT omjeri alela u uzorku su također pokazivali odstupanja od omjera 1:1, ali u manjoj mjeri i kod malog broja jedinki. Neki od mogućih razloga većih odstupanja su 1) neravnomjerna učestalost alela u uzorku zbog toga što se određeni alel slabije umnaža zbog slabijeg vezanja početnica ili 2) neravnomjerna raspodjela alela u 5 000 očitavanja koje obrađuje AmpliSAS. Budući da je u svim uzorcima s omjerom alela 1:3 manje zastupljen alel CacaDRB1*0301, a isti je također slabije zastupljen i kod ostalih jedinki koje ga imaju te je jedini alel s učestalošću manjom od 30% u bilo kojem uzorku u kojem je nađen, mogli bismo zaključiti da se slabije umnaža zbog slabijeg vezanja početnica. Razlog tome moglo biti postojanje jednog ili više polimorfizama u mjestima vezanja početnica na dotičnom alelu. Korištene početnice su dizajnirane za goveda (Sigurdardottir i sur., 1991), što je dodatni razlog eventualnog nepodudaranja.

Promjena parametara nije utjecala na konačne rezultate analize. Promjene učestalosti alela primijetila sam samo kod graničnih vrijednosti parametara grupiranja; vrlo malih - MDF manje od 1% ili vrlo velikih - postotak sličnosti veći od 99%, što je u skladu s očekivanim za vrste s jednim lokusom MHC kojeg istražujemo i kod kojih je lako odvojiti artefakte od pravih alela. Zbog toga nema potrebe za optimizacijom i u potpunosti je dovoljno koristiti samo zadane parametre grupiranja. Parametri filtriranja ne utječu na učestalosti sekvenci, već samo uklanjaju neke od rezultata na temelju prepoznavanja artefakta ili male opće zastupljenosti (parametar MAF). Za izbor parametara filtriranja mogu nam pomoći rezultati modula AmpliCHECK u kojem su prikazane učestalosti i vrsta artefakata (Tablica 6.)

Za razliku od promjene parametara, podjelom datoteke na dijelove kako bi se analiziralo što više različitih očitavanja dobila sam veće promjene učestalosti alela (Tablica 10), no očitavanja pojedinih alela su poprilično ravnomjerno zastupljena u cijeloj datoteci i ne mijenja se konačni rezultat – isti aleli su dominantni kada se analizira bilo koji dio datoteke te 5 000 očitavanja predstavlja dobru aproksimaciju i u ovom slučaju je potpuno dovoljno za pouzdanu genotipizaciju. Za vrste s udvostručenim lokusima MHC bilo bi korisno analizirati što više očitavanja i koristiti računalnu verziju programa AmpliSAT koja pruža dodatne mogućnosti analize kako bi se moglo s većom sigurnošću zaključiti je li određena sekvenca slabije zastupljen alel ili jače zastupljen artefakt.

Budući da uzorak u ovom radu nije bio nasumično odabran nije moguće procjenjivati genetičku raznolikost populacije. Od osam alela nađenih u ovom istraživanju njih pet, koji su ujedno i najzastupljeniji, je poznato i nalazi se u bazi podataka Genbank. Od alela koji se ne nalaze u Genbank bazi, aleli MHC-0001 i MHC-0002 nađeni su u neobjavljenom istraživanju slovenskih populacija srna (Bužan, neobjavljeni rezultati), a alel MHC-0003 potpuno je nov i potrebno ga je potvrditi daljnjim istraživanjima budući da se nalazi samo kod jedne heterozigotne jedinke. Aleli MHC-0001 i MHC-0003 se od najsličnijih alela razlikuju po svega jednom polimorfnom mjestu što upućuje na moguću nedavnu divergenciju alela, što bi trebalo potvrditi u daljnjim istraživanjima.

Tri alela (MHC-0001, CacaDRB1*0301 i Caca-DRB*0302) dugačka su 267 pb; sadrže deleciju kodona što je relativno rijedak nalaz kod MHC alela, koji se međusobno razlikuju većinom u supstitucijskim mutacijama. Deleciju istog kodona, kodona 65, također sadrže aleli DRB3 i DQB lokusa domaćeg goveda, *Bos taurus* (Linneaus, 1758), (BoLA-DRB3*0201, *3301, *1401 i DQB3*1). Razlog tome može biti mutacijski događaj kod zajedničkog pretka porodica goveda i jelena ili može biti posljedica konvergentne evolucije što dodatno podupire to što kod ostalih devet genotipiziranih vrsta iz porodica goveda i jelena nisu nađeni aleli s delecijom (Mikko i sur., 1997, Van den Bussche i sur 1999), a kod domaćeg goveda se alel BoLA-DRB3*0201 pojavljuje s velikom učestalošću u europskim populacijama. Kodon 65 kodira za aminokiselinu koja čini α zavojnicu i smatra se da delecija pruža funkcionalnu prednost u vezanju antigena jer je zavojnica kraća za jednu aminokiselinu „ravnija“ te sliči α zavojnici kakvu imaju receptori MHC skupine I (Mikko i sur, 1997).

Ostale mutacije su supstitucijske i broj nesinonimnih supstitucija je veći od sinonimnih (13 nesinonimnih nasuprot 10 sinonimnih) što je u skladu s očekivanjem jer je to jedna od karakteristika gena MHC, koji su pod utjecajem pozitivne selekcije (Van den Bussche, 1999).

6. ZAKLJUČAK

Priprema knjižnice fragmenata; umnoženog egzona 2 lokusa DRB 14 uzoraka srne, za sekvenciranje tehnologijom Ion Torrent bila je uspješna te je za svaki uzorak dobiven vrlo velik broj očitavanja visoke kvalitete.

Analizom dobivenih očitavanja u *online* programskom paketu AmpliSAT dobiveni su jasni i jednoznačni rezultati; promjenom parametara mijenjale su se učestalosti alela, no to nije utjecalo na konačnu determinaciju.

Od osam alela pronađenih u 14 uzoraka pet ih je od prije poznato te se nalazi u bazi podataka GenBank (CacaDRB1*0301, Caca-DRB*0201, *0302, *0304 *0401), dva, MHC-0001 i MHC-0002, se poklapaju s alelima nađenim u istraživanju slovenskih populacija (Bužan, neobjavljeni rezultati), a jedan, MHC-0003, je potpuno nov.

Za primjenu u genotipizaciji vrsta bez udvostručenih lokusa koje istražujemo i bez dugačkih homopolimernih regija sekvenciranje tehnologijom Ion Torrent daje pouzdane rezultate i za analizu je u potpunosti dovoljno koristiti zadane parametre *online* verzije programa AmpliSAT.

7. LITERATURA

Alberts B., Johnson A., Lewis J., Raff M., Roberts K., Walter P. (2008): *Molecular Biology of the Cell*, 5th edition, Garland Science, New York.

Babik W.(2010):Methods forMHCgenotyping in non-model vertebrates. *Molecular Ecology Resources*, 10:237–251.

Ballard D., Winkler-Galicki J., Wesoły J. (2020): Massiveparallelsequencinginforensics: advantages, issues, technicalities, andprospects.*International Journal of Legal Medicine*, 134; 1291–1303.

Baxevanis AD., and B. F. Francis Ouellette BF. (2001): *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins (Methods of Biochemical Analysis, 43)*, 2nd edition, John Wiley & Sons, New York.

Bishop GA.(1998): *Encyclopedia of Immunology* (second edition), Academic Press, Cambridge.

Glenn TC.(2011): Field guide to next-generation DNAsequencers. *Molecular Ecology Resources*, 11: 759–769.

Kelley J., Walter L., Trowsdale J.(2004): Comparative genomics of major histocompatibility complexes. *Immunogenetics*, 56: 683–695.

Kindt T., Goldsby R., Osborne B. (2007):. *Kuby immunology*. W. H. Freeman & Company, New York.

Kumanovics A., Takada T., Lindahl KT.(2003): Genomic organization of themammalianMHC. *Annual Review of Immunology*,21: 629–57.

Kumar KR., Cowley MJ., Davis RL.(2019): Next-generation sequencing and emerging Technologies. *Seminars in Thrombosis and Hemostasis*, 45(07): 661-673.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. (2009): The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25: 2078–2079.

- Mellins ED., Stern LJ. (2014): HLA-DM and HLA-DO, key regulators of MHC-II processing and presentation. *Current Opinion in Immunology*, 26: 115–22.
- Mikko S., Lewin HA., Andersson L. (1997): A phylogenetic analysis of cattle DRB3 alleles with a deletion of codon 65. *Immunogenetics*, 47: 23-29.
- Mikko S., Knut R., Schmutz S., Andersson L. (1999): Monomorphism and polymorphism at MHC DRB loci in domestic and wild ruminants. *Immunological Reviews*, 167: 169-178.
- Pareek CS., Smoczynski R., Tretyn A. (2011): Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, 52, 413–435.
- Perez-Espona S., Goodall-Copestake WP., Savirina A., Bobovikova J., Molina-Rubio C., Pérez-Barbería FJ. (2019): First assessment of MHC diversity in wild Scottish red deer populations. *European Journal of Wildlife Research*, 65: 22.
- Quemere E., Galan M., Cosson JF., Klein F., Aulagnier S., Gilot-Fromont E., Merlet J., Bonhomme M., Hewison AJM., Charbonnel N. (2015): Immunogenetic heterogeneity in a widespread ungulate: the European roe deer (*Capreolus capreolus*). *Molecular Ecology*, 24: 3873–3887.
- Sebastian A., Herdegen M., Migalska M., Radwan J. (2016): AMPLISAS: a web server for multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology Resources*, 16: 498–510.
- Sebastian A. (2018): AmpliSAT Manual; Amplicon Sequencing Analysis Tools, v. 8 (<http://evobiolab.biol.amu.edu.pl/amplisat/index.php?documentation>)
- Sommer, S. (2005): The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology*, 2(1), 16.
- Stipoljev S., Bužan E., Rolečková B., Iacolina L., Šprem N. (2020): MHC Genotyping by SSCP and Amplicon-Based NGS Approach in Chamois. *Animals*, 10: 1694.
- Ujvari B., Belov K. (2011): Major Histocompatibility Complex (MHC) Markers in Conservation Biology. *International Journal of Molecular Sciences*, 12: 5168-5186.

Van den Bussche R A., Hooper SR., Lochmiller RL.(1999): Characterization of MHC-DRB allelic diversity in white-tailed deer (*Odocoileus virginianus*) provides insight into MHC-DRB allelic evolution within Cervidae. *Immunogenetics*, 49: 429–437.

<https://www.thermofisher.com/order/catalog/product/A27763#/A27763> - pristupljeno 15.6.2021.

https://groups.google.com/g/amplisat/c/agbS2G2Piwg?fbclid=IwAR0SY4LVJHLrOnAshWVQnwX_CA1P-IcHmtbgUA4pKSfdxMwoRVaanXKTfHc - pristupljeno 21.11.2021.

https://www.thermofisher.com/blog/behindthebench/how-many-samples-can-i-multiplex-on-my-ngs-run-seq-it-out-13/?fbclid=IwAR1JepKTgHExEWfQuEwW0nzgFTNCDOglhBdFWeJOYD2gNSa_q1iTBXbT1nA - pristupljeno 20.11.2021.

Korišteni programi i stranice

AmpliSAT - <http://evobiolab.biol.amu.edu.pl/amplisat/>

Clustal Omega - <https://www.ebi.ac.uk/Tools/msa/clustalo/>

EMBOSS Needle - https://www.ebi.ac.uk/Tools/psa/emboss_needle/

ExPASy - <https://web.expasy.org/translate/>

GenBank® - www.ncbi.nlm.nih.gov/genbank/

JalView - <http://www.jalview.org/getdown/release/>

Notepad++ - <https://notepad-plus-plus.org/downloads/>

SAMtools - <https://github.com/samtools/>

8. PRILOZI

Prilog 1. Rezultati pročišćavanja datoteka fastq u modulu AmpliCLEAN.

| uzorak | broj sirovih očitanja | broj očitavanja nakon pročišćavanja | % uklonjenih očitanja |
|--------|--------------------------|---|-----------------------------|
| 1SN | 206 391 | 45 072 | 78,2 |
| 1SL | 213 718 | 48 228 | 77,4 |
| 2SL | 376 344 | 87 682 | 76,7 |
| 3SL | 278 302 | 62 332 | 77,6 |
| L2 | 310 701 | 67 470 | 78,3 |
| L5 | 264 817 | 61 115 | 76,9 |
| L19 | 292 534 | 70 654 | 75,8 |
| L20 | 211 862 | 22 466 | 89,4 |
| K7 | 332 978 | 120 181 | 63,9 |
| K10 | 204 747 | 43 384 | 78,8 |
| 4SC | 317 923 | 61 976 | 80,5 |
| 12SC | 360 473 | 63 571 | 78,2 |
| 13SC | 255 962 | 66 342 | 77,4 |
| 14SC | 195 903 | 39 316 | 76,7 |

9. ŽIVOTOPIS

Rođena sam 12.3.1997. u Zagrebu, gdje sam završila osnovnu školu i opći smjer gimnazije. Godine 2015. sam upisala preddiplomski studij biologije, a 2018. diplomski studij molekularne biologije na Prirodoslovno-matematičkom fakultetu Sveučilišta u Zagrebu. Tijekom studiranja razvila sam interes za mnoga područja biologije, ponajviše molekularnu biologiju i ekologiju, kojima bi se i dalje htjela baviti. Sudjelovala sam u nekoliko manjih izvanastavnih aktivnosti kao što su „Noć biologije“, „Mala škola mamalogije“ i „Zelena čistka“ u organizacijama Udruge studenata biologije (BIUS) i udruge BIOM. Laboratorijsku stručnu praksu u sklopu studiranja odradila sam na Institutu Ruđer Bošković, u laboratoriju za molekularnu genetiku, i na Zavodu za animalnu fiziologiju, u sklopu projekta „Čigra“ te sam volontirala na projektu „Promjene sastava patogena i imunološkog odgovora tijekom širenja areala uspješnih invazivnih vrsta slatkovodnih rakova“. Diplomski rad sam izradila u sklopu erasmus stručne prakse u laboratoriju za Bioraznolikost Fakulteta matematike, prirodoslovlja i informacijske tehnologije u Kopru gdje sam dodatno usavršila metode istraživanja DNA i vještine rada u laboratoriju.