

Usporedba metoda genotipizacije glavnog sustava tkivne podudarnosti u genomu s dupliciranim lokusom

Martinčić, Jelena

Master's thesis / Diplomski rad

2021

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://urn.nsk.hr/urn:nbn:hr:217:827073>

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

Download date / Datum preuzimanja: **2024-05-03**



Repository / Repozitorij:

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



Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Jelena Martinčić

**Usporedba metoda genotipizacije glavnog sustava
tkivne podudarnosti u genomu s dupliciranim lokusom**

Diplomski rad

Zagreb, 2021.

University of Zagreb
Faculty of Science
Department of Biology

Jelena Martinčić

**Comparison of MHC genotyping methods in the
genome with duplicated locus**

Master thesis

Zagreb, 2021.

Ovaj rad je izrađen u Laboratoriju za molekularnu ekologiju na Zavodu za animalnu fiziologiju Prirodoslovno-matematičkog fakulteta u Zagrebu, pod voditeljstvom izv. prof. dr. sc Ane Galov. Rad je predan na ocjenu Biološkom odsjeku Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu radi stjecanja zvanja magistar eksperimentalne biologije.

ZAHVALA

Zahvaljujem svojoj mentorici izv. prof. dr. sc. Ani Galov na izdvojenom vremenu, trudu i strpljenju bez kojeg ne bih uspjela završiti ovaj diplomski rad. Također, želim zahvaliti i asistentici mag. biol. exp. Idi Svetličić na svim savjetima, pomoći tijekom pisanja te samog laboratorijskog rada. Najljepše hvala i Gordani Žakman na tehničkoj podršci.

Veliko hvala mojim roditeljima Jasminki i Božidaru, sestri, šogoru te bakama i djedovima čija me podrška pratila kroz sve godine studiranja. Na bezuvjetnoj potpori, ohrabrenjima te razumijevanju zahvalnost želim izraziti i dečku Mariu. Hvala Vam na vjeri u mene!

Svim sadašnjim prijateljima, a i onima koji više nisu tu, također posebna hvala. Svi ste u nekom razdoblju bili dio mog fakultetskog života, a samim time i pridnonijeli njegovom završetku.

Na kraju, od srca zahvaljujem svojim životnim prijateljicama Valentini Horak te Eleni Jedvaj na ljubavi, strpljenju, razumijevanju kao i svim proživljenim zajedničkim trenucima.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Diplomski rad

Usporedba metoda genotipizacije glavnog sustava tkivne podudarnosti u genomu s dupliciranim lokusom

Jelena Martinčić

Rooseveltov trg 6, 10000 Zagreb, Hrvatska

Geni glavnog sustava tkivne podudarnosti (eng. *major histocompatibility complex*, MHC) kodiraju MHC molekule sustava I i II čija je funkcija predočavanje antigena pri pokretanju adaptivnog imunskog odgovora. Tijekom evolucije kod MHC gena događa se duplikacija lokusa koja povećava varijabilnost, a time i sposobnost obrane organizma. Duplikaciju lokusa karakterizira pojava više od dva alela po jedinci, koji se tijekom genotipizacije koamplificiraju s određenim parom početnica. Determinacija alela se, stoga, uz metodu Sangerovog sekvenciranja upotpunjuje metodom molekularnog kloniranja ili se genotipizacija temelji na metodama sekvenciranja nove generacije (eng. *next generation sequencing*, NGS). U ovom diplomskom radu usporedila sam tri metode genotipizacije MHC sustava na primjeru nemodelnog kraljevnjaka obični jelen *Cervus elaphus* (Linnaeus, 1758). Genotipizirala sam egzon 2 DRB lokusa MHC sustava II u 10 uzoraka vrste obični jelen te dobivene rezultate analizirala u računalnim programima Bioedit, Seqscape te računalnom paketu AmpliSAT. Kao najoptimalnija pokazala se metoda sekvenciranja zasnovana na sintezi (Illumina) kojom je identificirano najviše alela, njih 19. Sa sve tri metode identificiran je ukupno 21 alel, od čega 9 nije do sada bilo otkriveno.

(51 stranica, 6 slika, 4 tablica, 45 literaturnih navoda, jezik izvornika: hrvatski)

Rad je pohranjen u Središnjoj biološkoj knjižnici

Ključne riječi: obični jelen, multiplikacija, koamplifikacija, DRB lokus, NGS, Sangerovo sekvenciranje, AmpliSAT

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

Ocjenitelji: Izv. prof. dr. sc. Ana Galov

Izv. prof. dr. sc. Jasna Lajtner

Prof. dr. sc. Antun Alegro

Rad prihvaćen: 4.11.2021

BASIC DOCUMENTATION CARD

University of Zagreb
Faculty of Science
Department of Biology

Master Thesis

Comparison of MHC genotyping methods in the genome with duplicated locus

Jelena Martinčić

Rooseveltova trg 6, 10000 Zagreb, Hrvatska

The genes of the major histocompatibility complex (MHC) encode MHC class I and II receptors whose function is to present the antigen when triggering an adaptive immune response. During the evolution of the MHC genes, duplication of loci takes place, which increases the variability and thus the ability of the organism to defend itself. Duplication of loci is characterized by the appearance of more than two alleles per individual, which during genotyping co-amplify with a particular pair of primers. Along with the method of Sanger sequencing supplemented by the molecular cloning method, genotyping is based on next generation sequencing (NGS) methods. In this thesis, I compared three methods of genotyping the MHC system on the example of the non-model vertebrate red deer, *Cervus elaphus* (Linnaeus, 1758). I genotyped exon 2 of the DRB locus of the MHC system II in 10 samples of the red deer species and analyzed the obtained results in the computer programs Bioedit, Seqscape and the computer package AmpliSAT. The sequencing by synthesis (Illumina) NGS method, which identified the most alleles (19) proved to be the most optimal. A total of 21 alleles were identified with all three methods, of which 9 have not been detected so far.

(51 pages, 6 figures, 4 tables, 45 references, original in: croatian)

Thesis is deposited in Central Biological Library.

Keywords: red deer, multiplication, coamplification, DRB locus, NGS, Sanger sequencing, AmpliSAT

Supervisor: Assoc. Prof. Ana Galov

Reviewers: Assoc. Prof. Ana Galov

Assoc. Prof. Jasna Lajtner

Prof. Antun Alegro

Thesis accepted: 4.11.2021

POPIS KRATICA

IUCN - međunarodni savez za očuvanje prirode (eng. International Union for Conservation of Nature)

MHC - glavni sustav tkivne podudarnosti (eng. Major Histocompatibility Complex)

HLA - glavni sustav tkivne podudarnosti u čovjeka (eng. Human Leukocyte Antigens)

PCR - lančana reakcija polimerazom (eng. Polymerase Chain Reaction)

RFLP - polimorfizam duljine restrikcijskih fragmenata (eng. Restriction Fragment Length Polymorphism)

SSCP - (eng. Single Strand Conformation Polymorphism)

TBE - tris- borat - EDTA (eng. Tris-Borate-Ethylenediaminetetraacetic acid)

GLB - pufer za nanošenje uzoraka (eng. gel loading buffer)

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

UV - ultraljubičasta (eng. ultraviolet)

pb - parova baza

SADRŽAJ

1. UVOD.....	1
1.1. Geni glavnog sustava tkivne podudarnosti	1
1.1.2. Duplikacija lokusa na primjeru vrste obični jelen.....	4
1.2. Metode genotipizacije glavnog sustava tkivne podudarnosti u divljih vrsta životinja.....	5
1.2.1. Metoda Sangerovog sekvenciranja i molekularnog kloniranja.....	5
1.2.2. Metode sekvenciranja nove generacije.....	6
1.2.2.1. Metoda NGS zasnovana na sintezi (Illumina).....	7
1.2.2.2. Metoda NGS zasnovana na promjeni pH (Ion Torrent).....	12
1.3. Pregled dosadašnjih metoda genotipizacije MHC sustava kod običnog jelena.....	14
1.4. Cilj diplomskog rada.....	16
2.MATERIJAL I METODE	17
2.1. Uzorci tkiva jelena	17
2.2. Izolacija DNA.....	17
2.3. Lančana reakcija polimerazom.....	18
2.4. Elektorforeza u agaroznom gelu	19
2.5. Molekularno kloniranje	19
2.6. Sekvenciranje.....	23
2.9. Računalna analiza nukleotidnih sljedova.....	25
2.9.1. Analiza rezultata Sangerovog sekvenciranja u programima Bioedit i Seqscape.....	25
2.9.2. Analiza rezultata NGS sekvenciranja korištenjem bioinformatičkog alata AmpliSAT.....	26
3. REZULTATI.....	27
3.1. Sangerovo sekvenciranje i molekularno kloniranje.....	27
3.2. Metoda NGS-a zasnovana na sintezi (Illumina)	30

3.3. Metoda NGS-a zasnovana na promjeni pH (Ion Torrent)	32
3.4. Usporedba rezultata dobivenih različitim metodama genotipizacije	34
4. RASPRAVA.....	38
5. ZAKLJUČAK.....	44
6. LITERATURA	45
7. ŽIVOTOPIS	50
8. PRILOZI	51

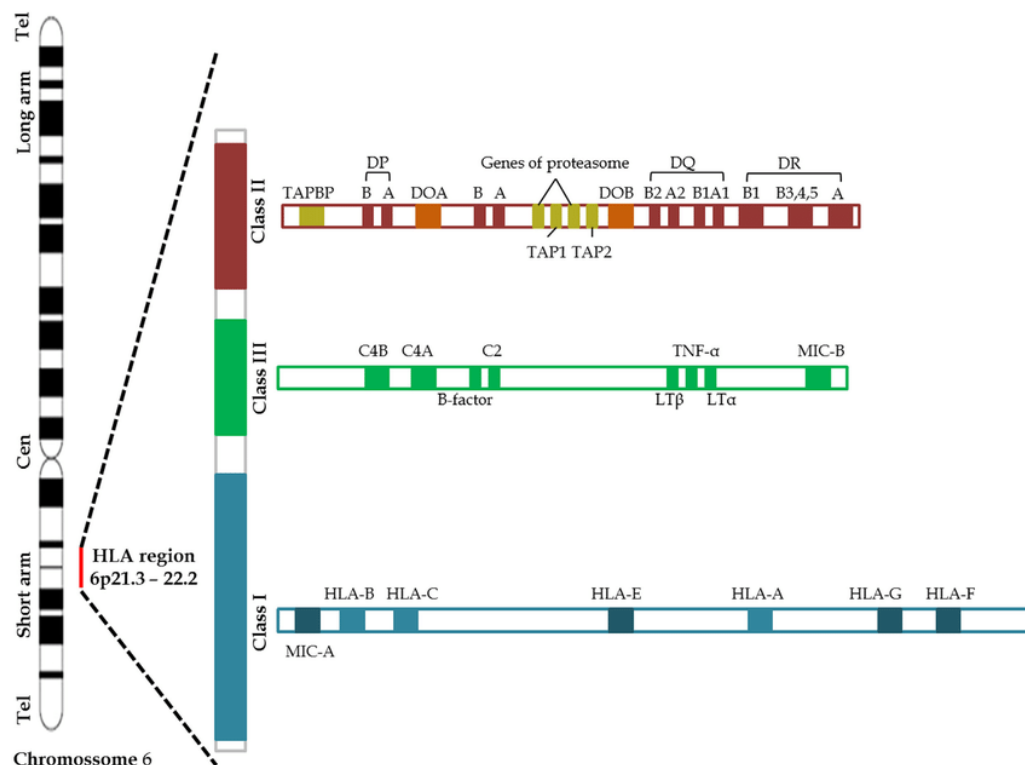
1. UVOD

1.1. Geni glavnog sustava tkivne podudarnosti

Rastući broj studija potvrđuje činjenicu kako genetska raznolikost populacije ima važnu ulogu u borbi protiv patogena te u razvoju epidemije određene zarazne bolesti (Sommer 2005). Među genima odgovornim za navedene karakteristike su geni glavnog sustava tkivne podudarnosti (MHC, eng. *major histocompatibility complex*). Osim u otpornosti, MHC geni imaju ulogu i u kondiciji, razvoju autoimunih bolesti, privlačenju partnera te reprodukciji, što direktno utječe i na preživljavanje jedinke (Shiina i sur. 2016). Kod ljudi dokazana je uloga u razvoju nuspojava na lijekove, u interakcijama između neuroloških stanica, sinaptičkim funkcijama, ponašanju te u neurološkim i psihijatrijskim poremećajima (Shiina i sur. 2016).

Pripadaju među najpolimorfnije gene do sada opisane u kralježnjaka. Seleksijski pritisak koji je doveo do očuvanja toliko različitog broja alela između jedinki još nije sasvim poznat. Izraziti polimorfizam održava se brojnim genetskim mehanizmima poput konverzija, rekombinacija, točkastih mutacija te duplikacija gena. U ljudi MHC geni se nalaze na kratkom kraku kromosoma šest (q21.3), gdje obuhvaćaju oko 3,5 megabaza te čine genski sustav definiran kao sustav glavnih antigena tkivne snošljivosti (HLA) razreda I i II (Slika 1.) (Abbas i sur. 2018).

Raznolikosti MHC molekula pojedine jedinke, uz polimorfizam, pridonosi kodominantna ekspresija tih gena, tj. izražavaju se svi aleli naslijeđeni od oba roditelja. Ako je jedinka heterozigot, tada se očekuje da se većina ili svi setovi alela razlikuju između naslijeđenih haplotipova. Također, MHC gene karakterizira i poligenija (Slika 1). Haplotip MHC skupine I sastoji se od lokusa A, B i C, dok se haplotip skupine II sastoji od lokusa DP, DQ, DR. S obzirom da je molekula MHC skupine II sastavljena od α i β lanaca, svaki lokus haplotipa sadrži gene (A i B) koji kodiraju za α , tj. β lance. Navedeni lanci mogu se povezivati čak i ukoliko su eksprimirani u različitim haplotipovima. Prema tome, ako je jedinka heterozigot, svaka njena stanica može eksprimirati do šest različitih MHC molekula skupine I te između 8 i 12 različitih molekula skupine II, što pridonosi učinkovitijoj obrani organizma (Abbas i sur. 2018).



Slika 1. Prikaz smještaja (kratki krak kromosoma šest q21.3) te poligenije MHC sustava.

Preuzeto iz Lima-Junior i Pratt-Riccio (2016).

S obzirom na funkciju, produkti ovih gena eksprimiraju tri skupine MHC molekula: MHC molekule skupine I, II i III. Središnju ulogu u pokretanju adaptivnog imunološkog odgovora prezentacijom antigena limfocitima T imaju homologne skupine I i II. Lokusi nepolimorfne skupine III kodiraju citokine, molekule uključene u obradu antigena, kao i druge čimbenike imunoreakcije. Obrazac izraženosti MHC molekula na stanicama i tkivima u vezi je s ulogama limfocita T. Naime, ako strani antigen nije predstavljen u sklopu MHC molekula skupine I i II, limfociti T ih neće prepoznati niti će doći do posljedičnog pokretanja imunoreakcije. Stoga, ekspresija MHC molekula skupine I na svim stanicama s jezgrom osigurava prepoznavanje unutarstaničnih patogena ili tumorskih antigena, pomoću citotoksičnih T limfocita (CD8+). Naspram njih, skupina II izražena je samo na limfocitima B, dendritičkim stanicama te makrofagima. Kompleks antigen – molekula MHC skupine II na specifičnim stanicama u limfnim organima aktivira pomoćničke T limfocite (CD4+). Oni imaju ulogu u aktivaciji drugih imunoloških stanica, poput makrofaga ili limfocita B, koje zatim proizvode antitijela s ciljem uklanjanja izvanstaničnih antigena (Abbas i sur. 2018).

MHC molekule skupine I građene su od jednog α lanca te s njim nekovalentno povezanog β mikroglobulina. Lanac α kodiran je genima MHC, dok su geni za β mikroglobulin smješteni na 15. kromosomu. Molekula se sastoji od četiri dijela: djela koji veže peptide, djela sličnog imunoglobulinu, hidrofobnog transmembranskog i citoplazmatskog dijela koji sidri molekulu u staničnu membranu. Pri tome α lanac formira tri izvanstanične domene (α_1 , α_2 , α_3) te transmembransku i citoplazmatsku domenu. Domene α_1 i α_2 , kodirane egzonima dva i tri, građene su svaka od 90 aminokiselinskih ostataka te su odgovorne za vezanje različitih prerađenih antigenskih peptida. Utor veznog mjesta čini osam antiparalelnih β -ploča te dvije paralelne α -uzvojnice. Krajevi pukotine zatvoreni su tako da se u nju može vezati antigenski peptid veličine 8-11 aminokiselina. Varijabilnost aminokiselinskih ostataka na dnu i stjenkama pukotine različitih molekula MHC omogućuje vezanje različitih peptida. Domena α_3 odgovara dijelu molekule sličnom imunoglobulinu koja omogućuje povezivanje citotoksičnih limfocita T za vrijeme predočavanja antigena. Povezivanju pridonosi i dio β mikroglobulina te α_2 domene. Mikroglobulin β nepolimorfan je te nekovalentno vezan na α_3 domenu. Konzerviran je i bitan za ekspresiju MHC molekule skupine I na staničnoj membrani (Abbas i sur. 2018).

Molekule MHC skupine II heterodimeri su građeni od nekovalentno povezanih polimorfni α i β lanca. Oba lanca kodirana su MHC genima. Kod ljudi lanac β varijabilniji je u odnosu na lanac α . Cijela se molekula, također, može podijeliti na četiri jednaka dijela, kao i MHC molekula skupine I. Lanci tvore domene α_1 , α_2 , β_1 , β_2 . Vezno mjesto za peptide formiraju domene α_1 i β_1 , kodirane egzonima dva, dok su domene α_2 i β_2 konstantne i važne za vezanje pomoćničkih limfocita T. Utor je sastavljen od četiri β -nabrane ploče na dnu i jedne α -uzvojnice u stijenci svake od domena. Krajevi pukotine su otvoreni, stoga ona može vezati peptid veličine i više od 30 aminokiselinskih ostataka.

Za stabilnu ekspresiju obje skupine molekula neophodna je ekspresija sve tri podjedinice heterotrimer: vezanog antigenskog peptida, α lanca i β lanca, tj. β mikroglobulina. Nestabilne molekule se razgrađuju, a samo one s vezanim peptidom te pravilno građene se ekspimiraju na staničnoj površini.

1.1.2. Duplikacija lokusa na primjeru vrste obični jelen

Obični jelen, *Cervus elaphus* (Linnaeus, 1758) široko je rasprostranjena vrsta divljači u Europi te nekim dijelovima Afrike, Azije i Sjeverne Amerike. U posljednjih 50 godina zabilježen je nagli porast broja jedinki usprkos lošem lovnom gospodarenju te fragmentaciji staništa. Zbog toga ovu vrstu IUNC svrstava u kategoriju „najmanje zabrinjavajućih“ (Burbaite i Csányi 2010; Lovari i sur. 2018). Usprkos trendu ukupnog porasta, navedena je vrsta izumrla u Albaniji, dok je u Grčkoj pred izumiranjem (Burbaite i Csányi 2010). Obični jelen pripada carstvu životinja (Animalia), koljenu svitkovaca (Chordata), razredu sisavaca (Mammalia), redu parnoprstaša (Cetartiodactyla), porodici jelena (Cervidae), te rodu jelena (*Cervus*) (Lovari i sur. 2018).

Uloge svih vrsta unutar porodice *Cervidae* protežu se kroz niz ekonomskih, socijalnih i ekoloških dobrobiti. Naime, osim što su turistička atrakcija, važan su izvor hrane za životinje i ljude. Utječu na održavanje strukture i sastava vegetacije. No, ako im gustoća populacije prijeđe kritičnu točku, sposobni su nanijeti ozbiljnu štetu pašnjacima, šumama i obradivim površinama. Stoga, nužno je pravilno upravljanje populacijama radi održavanja zdravlja te ravnoteže i kvalitete okoliša (Burbaite i Csányi 2010).

Demografske promjene na razini populacija dešavaju se pod utjecajem evolucije i djelovanjem čovjeka. Naime, sve oboljele jedinke, te one s lošom kondicijom, umiru zbog nemogućnosti prilagodbe na različite okolišne uvjete i patogene. Čovjek na upravljanje populacijama utječe selektivnim lovnim odstrijelom pri čemu može doći do gubitka genetičke raznolikosti što može povećati osjetljivost na bolesti (Fernandez-de-Mera i sur. 2009). Prema tome, nužno je poznavanje MHC raznolikosti populacije čija je procjena kod vrste obični jelen otežana zbog duplikacije pojedinih lokusa koja uzrokuje prisutnost do šest alela po jedinici (Bujanić i sur. 2020). Evolucijski je duplikacija lokusa bitna zbog stvaranja novih funkcija gena te razvoja genoma. Osim za MHC, dokazane su i duplikacije gena mikrosatelita, receptora za miris te imunoglobulina (Huang i sur. 2019). Takvi lokusi često imaju slične sekvence DNA koje se mogu koamplificirati s istim parom početnica, što rezultira netočnim analizama. Osim varijacija u broju alela, analize otežavaju i nul-aleli, pseudogeni te suvišak ili nedostatak heterozigota (Huang i sur. 2019). Do danas je razvijen širok spektar analiza nukleotidnih sljedova takvih sekvenci, poput RFLP-a (eng. *Restriction Fragment Length Polymorphism*), SSCP-a (eng. *Single Strand Conformation Polymorphism*), genotipizacije mikrosatelita, Sangerovog sekvenciranja te metoda

sekvenciranja nove generacije (Babik 2010). Iako su neke od navedenih metoda vrlo pouzdane i precizne, obrazac duplikacije gena i dalje predstavlja problem za analizu prilikom genotipizacije.

1.2. Metode genotipizacije glavnog sustava tkivne podudarnosti u divljih vrsta životinja

1.2.1. Metoda Sangerovog sekvenciranja i molekularnog kloniranja

Sangerovo sekvenciranje metoda je prve generacije sekvenciranja. Zajedno sa suradnicima, tehniku je 1977. godine razvio Frederick Sanger sa Sveučilišta Cambridge. Za to je dostignuće 1980. godine nagrađen Nobelovom nagradom za kemiju. Tehnika je široko usvojena zbog niske radioaktivnosti te visoke učinkovitosti. Prvi genomi sekvencirani ovom tehnologijom su genomi bakteriofaga *phiX174* duljine 5374 bp te bakteriofaga λ duljine 48501 bp (Kchouk i sur. 2017).

Metoda se temelji na dodatku smjese kemijski modificiranih nukleotida nazvanih dideoksinukleozid-trifosfati (dNTP) tijekom sinteze novog DNA lanca kemijski ili genetski modificiranom DNA polimerazom kojoj nedostaje 3'-5' egzonukleazna aktivnost (Walker i Lorsch 2013). Razlikujemo radioaktivno ili fluorescentno obilježene dideoksinukleozid-trifosfate za adenin (ddA), timin (ddT), gvanin (ddG) i citozin (ddC). Karakteristika ovih nukleotida jest da nakon ugradnje u lanac, koče daljnju sintezu, zbog čega dobivamo fragmente DNA molekule različitih duljina, ovisno o tome gdje se određeni dNTP ugradio. Kako bi ih se razlikovalo, fragmenti se odvajaju po veličini gel elektroforezom te vizualiziraju ovisno o načinu označavanja (Maxam i Gilbert 1992; Prober i sur. 1987).

Razvojem tehnologije Sangerovo sekvenciranje sve više nadjačavaju metode sekvenciranja novih generacija. Razlog tome su nedostaci, poput nemogućnosti sekvenciranja složenih genoma, cijena te dugotrajna izvedba metode (Kchouk i sur. 2017). Također, kod sekvenciranja genoma jedinki s dupliciranim lokusima te heterozigotnih jedinki, u ispisu elektroferograma, dolazi do pojave višestrukih vrhova u sekvencama, što otežava genotipizaciju. Dakle, ono se često provodi uz metodu molekularnog kloniranja željenog odsječka DNA u bakterije *E. coli* s ciljem odvajanja alelnih faza. Molekularno kloniranje uvriježen je postupak koji je često neophodan kod genotipizacije dupliciranih lokusa. Konvencionalno, metoda se sastoji od protokola za izolaciju, umnažanje te ligaciju željenog DNA odsječka u vektor za kloniranje. Nakon ligacije odvija se transformacija rekombinantnih plazmida u odgovarajućeg domaćina (*E. coli*) te selekcija onih domaćina koji su usvojili plazmid s insertom. Svaka kolonija sadrži jedan produkt lančane reakcije

polimerazom (PCR, eng. *polymerase chain reaction*), tj. jedan odsječak DNA. Ako je selekcija uspješna, slijedi purifikacija te kako bi analizirali nukleotidne sljedove - metoda Sangerovog sekvenciranja. Osim što je za otkrivanje svih alela potrebno sekvencirati puno bakterijskih klonova, metoda molekularnog kloniranja sklona je i umnožavanju artefakata u bakterijama nastalih PCR-om, što utječe na točnost obrade podataka. Dodatni artefakti mogu se uvesti i sustavom popravka DNA domaćina koji se aktivira tijekom replikacije. Zbog svega navedenog, metode Sangerovog sekvenciranja i molekularnog kloniranja izbjegavaju se u studijama s velikim brojem uzoraka, gdje se često koriste radi dobivanja preliminarnih podataka zajedno s drugim metodama. Konkretno, u istraživanju MHC sustava bitne su za početnu procjenu sadržaja te varijaciju alela (Babik 2010; Fuselli i sur. 2018; Lighten i sur. 2014).

1.2.2. Metode sekvenciranja nove generacije

S obzirom na nedostatke metode prve generacije, 2004. godine komercijalno postaju dostupne metode sekvencera druge generacije. Tu pripadaju platforme koje provode sekvenciranje zasnovano na sintezi (*Sequencing by synthesis*) te ligaciji (*Sequencing by ligation*). Platforme Illumina genome analyzer iz 2006. godine te Ion Torrent iz 2010., iako tijekom reakcije sekvenciranja koriste različitu kemijsku podlogu, zbog čega platformu Ion Torrent zovemo još i metodom zasnovanom na promjenu pH, generalno pripadaju u tip tehnologije koja se temelji na sekvenciranju zasnovanom na sintezi (*Sequencing by synthesis*). Osim njih, kasnije su razvijene metode sekvenciranja treće i četvrte generacije zasnovane prvenstveno na razlikama u kemiji sekvenciranja, dužini formiranih očitavanja, stopama i vrstama pogrešaka, pokrivenosti genoma, cijeni te vremenu provedbe postupka. Zbog simultanog izvođenja umnožavanja i sekvenciranja milijuna fragmenata, ove se tehnologije nazivaju još i metodama masivnog paralelnog sekvenciranja (Ambardar i sur. 2016).

Sekvenciranje amplikona metodama sekvenciranja nove generacije (eng. *amplicon sequencing*) temelji se na ciljanom sekvenciranju dijelova genoma ili lokusa, što omogućuje detaljniji uvid u specifične gene ili regije od interesa. Naspram Sangerovog sekvenciranja, omogućuju sekvenciranje tisuće amplikona po reakciji u kraćem vremenu obrade, što ih čini isplativijim metodama. Osim toga, pružaju i veću osjetljivost u područjima genoma koje je teško sekvencirati, poput onih bogatih GC bazama. Velik broj dostupnih komercijalnih kompleta te protokola omogućuje veću fleksibilnost pri dizajnu eksperimenata, ovisno o vrsti te regijama od interesa.

Široku primjenu pronašle su u dijagnostici tumora, metagenomici, forenzici, genomici mikroba i slično (<https://www.illumina.com/techniques/sequencing/dna-sequencing/targeted-resequencing/amplicon-sequencing.html>, Galan i sur. 2010).

1.2.2.1. Metoda NGS zasnovana na sintezi (Illumina)

Sekvenciranje zasnovano na sintezi vrlo je pogodno za degradirane i uzorke s malom koncentracijom DNA molekula. Tijek rada slijedi četiri temeljna koraka: pripreme knjižnice, formiranja klastera, sekvenciranja i analize podataka. Za korak pripreme knjižnice može se upotrijebiti bilo koji od dostupnih komercijalnih paketa (Illumina). Generalno, priprema knjižnice obuhvaća: fragmentaciju uzorka (ili umnažanje ciljanog odsječka DNA pomoću PCR reakcije), popravak krajeva DNA molekule, dodatak adenina na 3' kraj te naposljetku adaptera. Nakon izolacije DNA se enzimatski ili fizički fragmentira, pri čemu se formiraju odsječci duljine od 300 do 500 pb ili se provede PCR reakcija s početnicama specifičnim za umnažanje fragmenta od interesa, a na koje su dodane sekvence indeksa (koje su specifične, odnosno različite za svaki uzorak, a prema kojima kasnije u analizi prepoznajemo pojedine uzorke). Kod duljih odsječaka uočena je manja učinkovitost metode. Nakon fragmentacije slijedi popravak krajeva DNA kako bi se postigli što optimalniji uvjeti za dodatak adaptera. Popravak obuhvaća formiranje ujednačenih tupih krajeva te fosforilaciju 5' krajeva. Učinkovitim povezivanju adaptera pridonosi još i dodatak adenina na sve 3' krajeve DNA molekule. Uloga adaptera vezana je uz imobilizaciju i umnožavanje fragmenta u kanalima protočnog polja čipa. Dodaju se dva jedinstvena, ali različita adaptera sastavljena od više povezanih manjih dodataka vidljivih na Slici 2.



Slika 2. Sivom bojom u sredini označen je DNA insert (fragment dvolančane DNA molekule originalnog uzorka). U plavim okvirima, označenim oznakama P5 adapter te P7 adapter, prikazane su podjedinice adaptera: crvenom bojom označene su oligonukleotidne sekvence P5 i P7 koje omogućavaju hibridizaciju za protočno polje čipa, narančastom bojom označeni su dijelovi adaptera Rd1 SP (*Read1 sequencing primer binding site - forward read*) i Rd2 SP (*Read2 sequencing primer binding site - reverse read*) koji čine vezno mjesto za početnice za sekvenciranje, crnom bojom označeni su indeksi 1 i 2 bitni za identifikaciju uzoraka prilikom sekvenciranja više uzoraka odjednom.

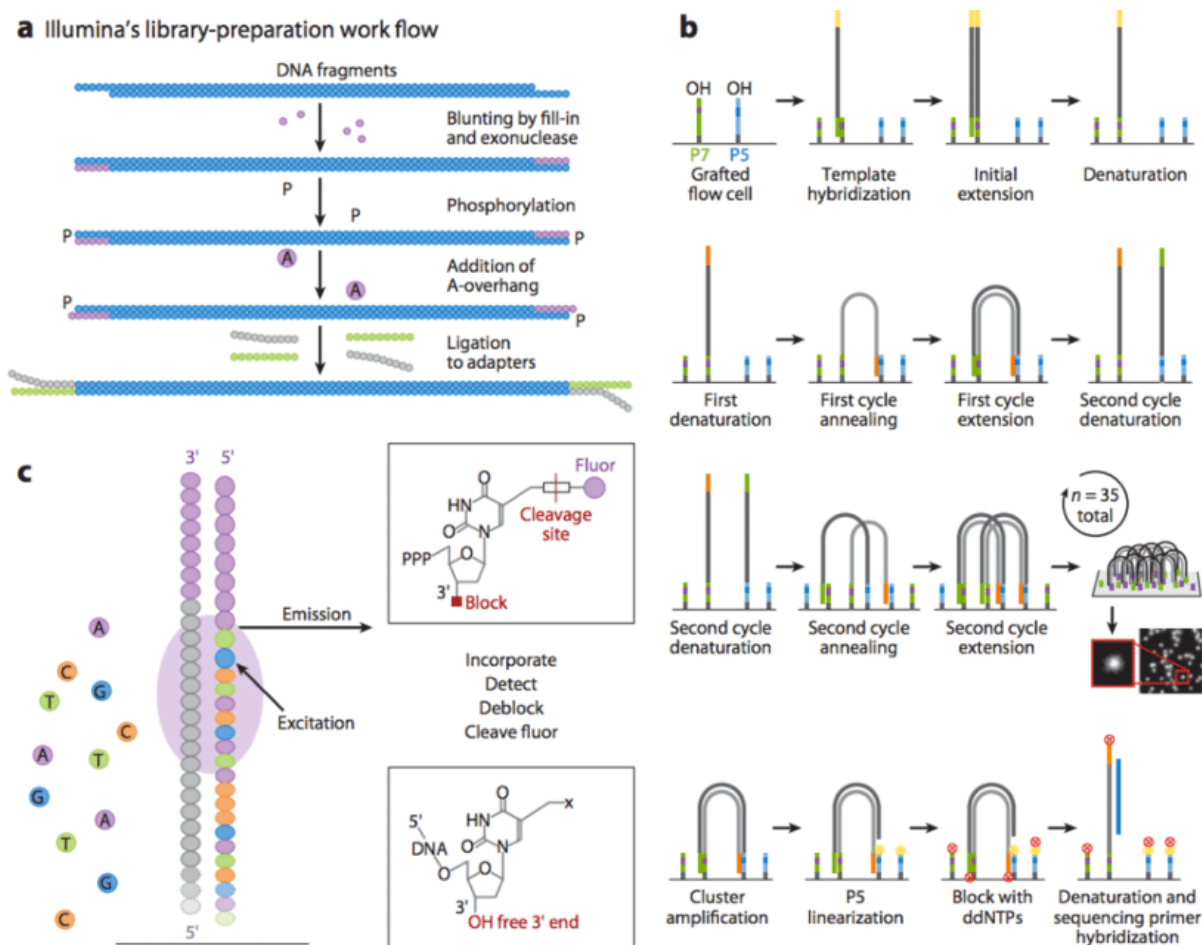
Preuzeto: <https://support.illumina.com/bulletins/2020/06/illumina-adapter-portfolio.html>

Formirana se knjižnica podvrgava još i pročišćavanju te koraku odabira samo onih odsječaka koji su odgovarajuće duljine. Svi odsječci koji su nepravilno vezali adaptore ili ih nisu vezali, odbacuju se. Nakon odabira odsječaka odgovarajuće duljine knjižnica se umnaža PCR-om s ciljem formiranja što veće količine DNA odsječaka koji bi mogli dati što točnije i kvalitetnije rezultate. Korak pripreme knjižnice prikazan je na Slici 3.a (Ambardar i sur. 2016; Bentley i sur. 2008; Bronner i sur. 2013).

Nakon pripreme knjižnice slijedi korak stvaranja nakupina (eng. clustering) koje podrazumijeva izotermalnu amplifikaciju svakog fragmenta nakon hibridizacije na oligonukleotide na površini protočnih polja pomoću adaptera dodanih u prvome koraku. Svaku nakupinu čine tisuće kopija jednog DNA odsječaka na nekoliko mikrona protočnog polja čipa. Uloga stvaranja nakupina je pojačanje signala prilikom sekvenciranja kako bi ono bilo što preciznije izvedeno. Izotermalnu amplifikaciju svakog fragmenta izvodi DNA polimeraza, nakon čega slijedi denaturacija dvolančane molekule DNA i ispiranje originalnog lanca. Svaki se amplificirani fragment potom savije poput mosta i hibridizira pomoću dodanog adaptera na drugi tip oligonukleotida. DNA polimeraza umnoži hibridizirani fragment formirajući „dvolančani most“. Sljedeći korak je

denaturacija koja rezultira dvjema jednolančanim kopijama originalnog lanca suprotne orijentacije vezanim za oligonukleotide na površini protočnog polja (Slika 3.b). Ovaj korak klonalnog umnožavanja ponavlja se kroz nekoliko ciklusa te se naziva *bridge amplification*, a rezultira simultanim nastankom milijuna kopija jednolančanih nakupina (Bentley i sur. 2008).

Sljedeći korak je sekvenciranje koje započinje uklanjanjem i ispiranjem jednog od lanaca DNA kako bi se dobili jednolančani identično orijentirani lanci. 3' krajevi molekule su blokirani, kako bi se spriječilo neželjeno formiranje dvolančanih mostova tijekom reakcije sekvenciranja. Sekvenciranje počinje produljivanjem početnice dodatkom smjese četiri fluorescentno obilježena nukleotida s terminatorom (2-cijanoetil) koji dopušta ugradnju samo jednog nukleotida u jednom ciklusu. Nakon ekscitacije svjetlom lasera svaki dodani nukleotid u svakom klasteru emitira karakteristični fluorescentni signal koji se potom bilježi optičkim čitačem. Svaka baza određena je specifičnom valnom duljinom. Nakon određivanja baze terminator i fluorescentni signal nukleotida se cijepaju i ispiru te počinje novi ciklus sekvenciranja (Slika 3.c). Kako bi se postotak greške održao ispod 1 %, izvodi se oko 150 ciklusa sekvenciranja. Za svaki se ciklus bilježe četiri uzastopne slike (Ambardar i sur. 2016; Kircher i sur. 2011).



Slika 3. Reakcija sekvenciranja zasnovana na sintezi (Illumina) (a) Slijed koraka pripreme knjižnice, (b) formiranja nakupina, (c) te reakcije sekvenciranja sintezom. Preuzeto iz Mardis (2013).

Duljina dobivene sekvence ovisi o broju provedenih ciklusa. Za svaku se nakupinu sve identične sekvence čitaju simultano. Nakon završetka prvog očitavanja formirana sekvenca se ispire i uklanja te se vrši sekvenciranje DNA lanaca iz suprotnog nizvodnog smjera (eng. PE, *paired – end sequencing*). Kako bi DNA polimeraza ponovno mogla formirati „dvolančani most“, odnosno jednolančani odsječak suprotne uzvodne orijentacije, blokada s 3' kraja se uklanja. Ponovno slijedi klonalno umnožavanje lanaca te uklanjanje uzvodnih lanaca čije je sekvenciranje provedeno u prethodnom koraku. Na formiranom jednolančanom komplementu ponovno se blokira 3' kraj te se vrši sekvenciranje iz suprotnog nizvodnog smjera. Tijekom reakcije sekvenciranja vrši se i sekvenciranje indeksa dodanih tijekom pripreme knjižnice. Indeksiranje je posebno korisno kod identifikacije svakog pojedinog uzorka kada se istovremeno sekvencira više različitih uzoraka.

Svaki je indeks veličine 6 do 10 bp te pruža jedinstvenu oznaku za svaki uzorak. Navedena procedura istovremenog sekvenciranja više uzoraka naziva se još i multipleksiranje (*multiplexing*).

Cijeli ovaj proces rezultira stvaranjem milijuna očitavanja DNA koji potiču iz svih fragmenata DNA sadržanih u uzorku. U četvrtom koraku analize podataka sekvence se razdvajaju prema jedinstvenim indeksima uvedenim u prvom koraku pripreme uzoraka. Računalni program određuje lokaciju svake nakupine te obrađuje i razdvaja za svaki pojedini ciklus sve zabilježene intenzitete (Kircher i sur. 2011). Svaki signal određenog intenziteta računalno se pretvara u određenu bazu koja odgovara tome intenzitetu signala. Za svaki uzorak, potom, dolazi do formiranja sparenih sekvenci (eng. contigs) sparivanjem očitavanja iz oba smjera (PE). Sparivanje očitavanja iz oba smjera omogućuje veću pokrivenost, veći broj očitavanja i podataka u usporedbi sa sustavima koji sekvenciraju fragmente samo iz jednog smjera. Također, u poravnanjima s referentnom knjižnicom pruža visokokvalitetne podatke, olakšava otkrivanje insercija i delecija te popunjavanje praznina u konsenzus slijedu formiranjem dužih konsenzus sekvenci (Ambardar i sur. 2016).

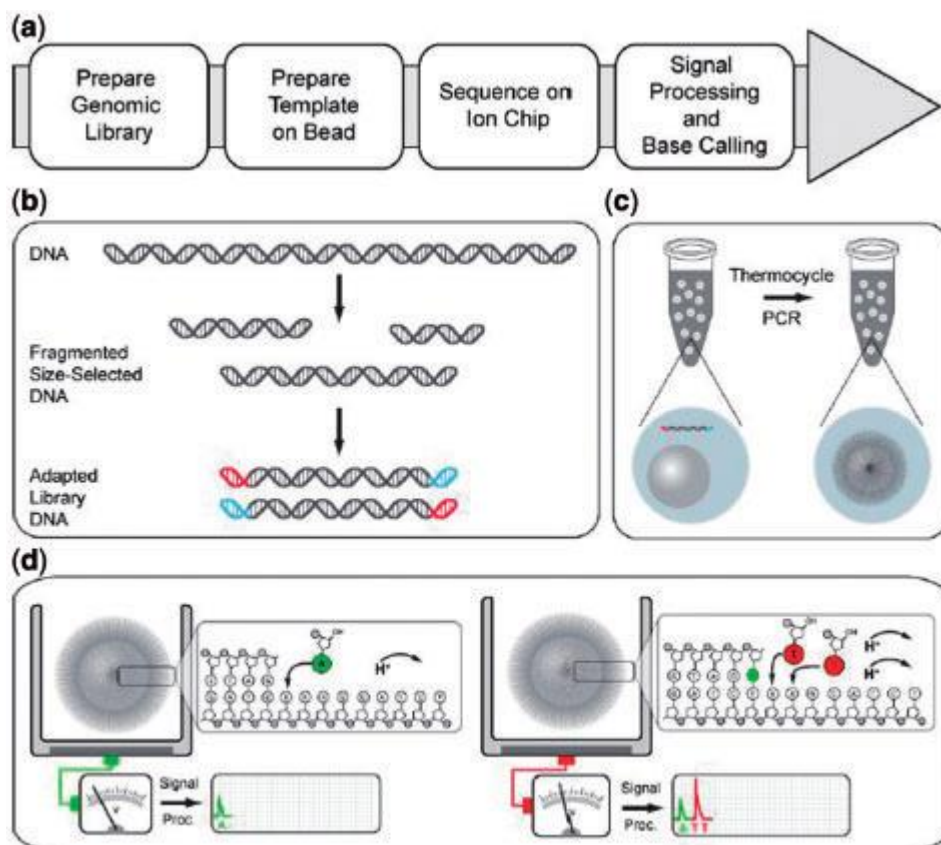
Najčešći tip pogreške koji prati ovu platformu jest supstitucija. Pogreške kod sekvenciranja homopolimera ovom metodom smanjene su zbog ugradnje samo jednog nukleotida u jednom ciklusu koji se za sljedeći mora ukloniti i isprati (Ambardar i sur. 2016; Reuter i sur. 2015).

1.2.2.2. Metoda NGS zasnovana na promjeni pH (Ion Torrent)

Tehnologija Ion Torrent temelji se na identifikaciji ugrađenog nukleotida tijekom reakcije sinteze novog lanca DNA zbog promjene pH uslijed oslobađanja protona. Platforma je zasnovana na četiri temeljna koraka: pripreme knjižnice, emulzijskog PCR-a, ugradnje u utor čipa i obrade signala. Proces započinje fragmentacijom DNA u mnoštvo fragmenata (ili umnažanjem ciljanog odsječka DNA pomoću PCR) koji, nakon što vežu adaptersku sekvencu, ulaze u mikroskopske „kuglice“ na način da jedan fragment ulazi u jednu kuglicu. Potom slijedi amplifikacija svakog pojedinog fragmenta DNA u svakoj kuglici emulzijskim PCR-om. Vezanje fragmenta DNA na kuglicu omogućuju ranije dodane adaptorske sekvence. Svaka kuglica predstavlja mikroreaktor u kojem se odvija umnažanje specifičnog fragmenta rezultirajući tisućama identičnih kopija koje su pričvršćene za svaku kuglicu. Ovaj korak je bitan jer dovodi do pojačavanja signala tijekom sekvenciranja. Naime, kada bi bio sekvenciran samo jedan lanac fragmenta, otpušteni proton ne bi davao dovoljno jak signal za bilježenje i mjerenje. Sljedeći korak jest ugradnja mikrokuglica u mikro-utore čipa, pri čemu svaki utor može primiti samo jednu mikrokuglicu. Jedan čip sadrži milijune utora (Reuter i sur. 2015; Rothberg i sur. 2011).

Proces sekvenciranja temelji se na oplahivanju bazenčića čipa, na čije je stijenke vezana DNA polimeraza zajedno s početnicama, s jednim od četiri tipa nukleotida. Svaki puta kada se određeni tip nukleotida inkorporira u rastući lanac, dolazi do hidrolize i oslobađanja jednog protona koji mijenja pH otopine u bazenčiću čipa. Promjena pH vrijednosti otopine proporcionalna je broju ugrađenih nukleotida (0,02 pH jedinice po jednom nukleotidu) (Rothberg i sur. 2011). Detekciju i konverziju promijenjenih pH vrijednosti u promjene u električnom naponu provodi osjetljivi sloj ispod bazena, konkretno ionski senzor ISFET (*ion sensitive field-effect transistor*). S obzirom na intenzitet promjene napona, bilježi se koliko je baza ugrađeno te formira slijed nukleotida sekvence. Ako se u otopini ne nalazi komplementarna baza lancu kalupa, nema oslobađanja protona ni promjene u naponu, te se ne bilježi ugradnja nijedne baze. U slučaju da se dvije identične baze nalaze jedna do druge, dolazi do inkorporacije dva nukleotida, te se voltaža udvostručuje, što se registrira kao detekcija dvije baze (<https://www.youtube.com/watch?v=zBPKj0mMcDg>, Rothberg i sur. 2011).

Varijacije u naponu bilježe se ispod svakog od milijun bazena u senzoričkom čipu pri čemu svaki sloj funkcionira kao maleni pH metar. Priprema knjižnice te reakcija sekvenciranja prikazani su na Slici 4.



Slika 4. (a) Tijek rada tehnologije Ion Torrenta, (b) priprema knjižnice fragmentacijom, odabirom optimalne duljine DNA odsječka te ligacijom adaptera, (c) klonalno umnažanje svakog odsječka u mikroskopskoj kuglici, (d) provođenje reakcije sekvenciranja u bazenčiću mikročipa.

Preuzeto iz Golan i Medvedev (2013).

Proces sekvenciranja ponavlja se svakih 15 sekundi, pri čemu se bazeni prethodno ispiru od nukleotida dodanih u prošlom ciklusu. Smanjena preciznost primijećena je kod sekvenciranja regija bogatih AT parovima te sekvenci koje imaju homopolimere duže od osam nukleotida zbog umnožavanja signala uslijed otpuštanja više protona. Prednost metode su veća dubina i pokrivenost naspram Sangerovog sekvenciranja te manja cijena zbog korištenja nemodificiranih baza (Ambardar i sur. 2016; Liu i sur. 2012).

1.3. Pregled dosadašnjih metoda genotipizacije MHC sustava kod običnog jelena

Usprkos dostupnosti velikog broja metoda, do danas je objavljeno tek nekoliko publikacija o genotipizaciji MHC sustava kod široko rasprostranjene vrste obični jelen *Cervus elaphus* (Linnaeus, 1758). Vodeći razlog koji otežava genotipizaciju kod ove nemodelne vrste organizma jest već spomenuti obrazac duplikacije gena koji se prilikom PCR reakcije koamplificiraju s određenim parom početnica (pogledati poglavlje 1.1.2.) (Babik 2010).

Među prvim istraživanjima provedenim u svrhu genotipizacije MHC sustava kod običnog jelena bile su studije provedene 1995. i 1997. godine (Swarbrick i sur. 1995; Swarbrick i Crawford 1997). U obje publikacije korištene su metode molekularnog kloniranja i Sangerovog sekvenciranja. Istraživanja su obuhvaćala srodno nepovezane jedinke, a Sangerovo sekvenciranje kloniranih sekvenci otkrilo je sekvence vrlo sličnog nukleotidnog sastava, poput onih kod ljudi i goveda. Po jedinci prosječno je sekvencirano 15 klonova te su pronađena od 1 do 4 različita alela, dok su u 50 analiziranih jedinki pronađena 34 alela. Osim prisutnosti više od jednog DRB i DQB lokusa, obje studije potvrdile su izraziti nukleotidni polimorfizam u djelu MHC molekule zaduženom za vezanje peptida. Kao vodeći nedostatak korištenih metoda navodi se niska vjerojatnost odabira i sekvenciranja onih klonova koji nose nisko zastupljene alele u populaciji (Swarbrick i sur. 1995).

Modificiranom metodom SSCP, Sangerovim sekvenciranjem te genotipizacijom mikrosatelita, 2009. godine provedena je karakterizacija alelne raznolikosti egzona 2 DRB lokusa podvrste jelena *Cervus elaphus hispanicus* (Hilzheimer, 1909)(Fernandez-de-Mera i sur. 2009; Fernández-De-Mera i sur. 2009). U studiji Fernandez-de-Mera i suradnika iz 2009. godine, osim izrazitog polimorfizma DRB lokusa, potvrđena je i povezanost određenih haplotipova s otpornošću na zarazne bolesti i parazite, veličinom slezene te kondicijskim stanjem životinja. Postavljena je hipoteza prema kojoj se navodi da određena alelna konfiguracija daje imunološku prednost, zbog čega su životinje rezistentnije na specifične patogene, a samim time imaju i veću slezenu i bolju kondicijsku spremu. U 94 jedinke pronađeno je 18 alela (od 1 do 4 različite alelne varijante po jedinci). Studija autora Fernández-De-Mera i suradnika iz 2009. godine također je potvrdila izrazit alelni polimorfizam, ali i dugoročni gubitak alelne raznolikosti kao posljedicu izolacije populacije, parenja u srodstvu (eng. inbreeding) te lova. Metodom SSCP te Sangerovim sekvenciranjem u 94 jedinke pronađeno je 18 različitih alela (od 1 do 4 različita alela po jedinci) te je u populaciji potvrđena ekspresija najmanje dva DRB lokusa (Fernández-De-Mera i sur. 2009).

Posljednje publikacije temelje se na metodama sekvenciranja nove generacije. Studija iz 2016. godine metodom sekvenciranja zasnovanom na sintezi (Illumina) potvrdila je pozitivnu korelaciju između određenih MHC genotipova, infekcije parazitima te razvoja rogovlja. U 155 ispitivanih jedinki pronađeno je 46 alelnih varijanti te od 1 do 4 različita alela po jedinci. Također, uočeno je da određeni genotipovi povećavaju otpornost za razvoj infekcije, dok pojava infekcije nije direktno povezana sa smanjenim razvojem rogovlja. Genotipovi koji omogućavaju otpornost na jednu vrstu parazita mogu istodobno uzrokovati povećanu osjetljivost na drugu vrstu (Buczek i sur. 2016).

Kako bi se potvrdili istinski aleli te oni zastupljeni u nižim frekvencijama, Pérez-Espona i suradnici 2019. godine proveli su genotipizaciju egzona 2 DRB lokusa običnog jelena u Škotskoj upotrebom metode Sangerovog sekvenciranja te metode sekvenciranja nove generacije Roche 454. U 48 analiziranih jedinki pronađeno je 25 alela. Svi aleli potvrđeni su s obje metodologije, no tri istinska alela pronađena su samo u po jednom klonu, zbog čega bi ih pristup Sangerovog sekvenciranja ispustio jer je za potvrdu istinskog alela bilo potrebno određeni alel pronaći u najmanje tri klona. Broj alela po jedinci kretao se od 1 do 6, što sugerira postojanje tri DRB lokusa u populaciji (Pérez-Espona i sur. 2019).

Metodom zasnovanoj na promjeni pH (Ion Torrent) analizirana je varijabilnost egzona 2 DRB lokusa 23 obična jelena (*Cervus elaphus*, Linnaeus, 1758) iz planinskog područja Republike Hrvatske. To je ujedno i prva studija na našem području, a broj pronađenih alela kretao se između 1 i 6, pri čemu je samo kod jedne jedinke pronađen jedan alel te kod jedne jedinke svih šest alelnih varijanti. Navedeno sugerira na postojanje najmanje tri lokusa u populaciji. Činjenica da je u dosta maloj veličini uzorka (23 životinje) pronađen velik broj alela (23 alela) ukazuje da su metode sekvenciranja nove generacije osjetljivije kod identifikacije različitih alela dupliciranih lokusa naspram npr. metode SSCP (Bujanić i sur. 2020).

1.4. Cilj diplomskog rada

Cilj ovog istraživanja jest usporedba triju metoda genotipizacije MHC sustava, na primjeru genotipizacije DRB lokusa MHC skupine II u vrste obični jelen (*Cervus elaphus* Linnaeus, 1758) koju karakterizira duplikacija DRB lokusa, zbog čega se u barem dijelu jedinki očekuje identifikacija više od dva alela po jedinci. Cilj je identificirati alele egzona 2 lokusa DRB u pojedinim uzorcima tkiva jelena korištenjem triju različitih metoda: I. metode Sangerovog sekvenciranja i molekularnog kloniranja, II. metode NGS zasnovanoj na sintezi (Illumina), te III. metode NGS zasnovanoj na promjeni pH (Ion Torrent). Krajnji cilj ovog diplomskog rada jest usporedbom utvrditi prikladnost pojedinih metoda za genotipizaciju MHC sustava u divljih vrsta životinja, pogotovo onih koje potencijalno imaju višestruke lokuse, odnosno kod kojih dolazi do koamplifikacije više lokusa s određenim parom početnica prilikom lančane reakcije polimerazom.

2.MATERIJAL I METODE

2.1. Uzorci tkiva jelena

Za istraživanje sam koristila 10 uzoraka jetre jedinki jelena prikupljenih lovnim odstrijelom na teritoriju Republike Hrvatske. Uzorci su prikupljeni 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 u razdoblju od 2016. do 2019. godine. Nakon uzimanja tkivo je pohranjeno na -20 °C u 96 % alkoholu u označenim plastičnim tubicama do daljnje analize.

2.2. Izolacija DNA

Iz pohranjenog tkiva izolaciju DNA izvršila sam pomoću prilagođenog protokola komercijalnog kompleta „Wizard Genomic DNA Purification Kita“ (Promega). Protokol se provodi u dva dana, pri čemu prvi završava inkubacijom preko noći na 55 °C uz povremeno miješanje na tresilici. U prvom sam koraku u označene tubice volumena 1,5 mL dodala 300 µL otopine za lizu stanica (Nuclei Lysis Solution). Nakon toga uzorke tkiva usitnila sam skalpelom te ih prema oznaci stavila u svaku tubicu. U trajanju od nekoliko desetaka sekundi, uzorke sam potom centrifugirala na 14 000 okretaja u minuti. Po završetku, uz svaku stjenku tubice, dodala sam 1,5 µL proteinaze K (20 mg/ml) koje sam, nakon ručnog protresanja, vorteksirala 5 minuta.

Drugi dan, prema protokolu, uzorke sam izložila sobnoj temperaturi te ih potom centrifugirala 10 sekundi s ciljem spuštanja komadića nerazgrađenog tkiva na dno tubice. Zatim je uslijedilo dodavanje 100 µL otopine za taloženje proteina (Protein Precipitation Solution) nakon kojeg sam tubice vorteksirala 20 sekundi. Nakon vorteksiranja tubice sam stavila na led (4 °C) 5 minuta. Sljedeći korak bio je centrifugiranje 3 minute na 13 000 okretaja u minuti. U talogu bijele boje zaostali su proteini, dok se DNA nalazila u supernatantu, tj. nadtalogu koji sam prebacila u nove tubice od 1,5 mL u koje sam prije dodala 300 µL 100 % etanola. Sadržaj tubice sam promiješala pažljivim okretanjem te potom uzorke centrifugirala 1 minutu na 13 000 okretaja u minuti. U ovom koraku DNA je zaostala u talogu, dok sam supernatant bacila. U tubicu s talogom dodala sam 300 µL 70 % etanola te ponovno lagano promiješala i centrifugirala 1 minutu na 13 000 okretaja u minuti. Nakon toga supernatant sam odstranila s pipetom. Zatim sam tubice s

talogom sušila 60 minuta na sobnoj temperaturi na filter papiru. Nakon sušenja dodala sam 100 μ L DNA otopine za rehidraciju (Rehydration Solution) te inkubirala na 4 °C. Izolirana DNA pohranjuje se u hladnjaku na 4 °C do sljedeće analize.

2.3. Lančana reakcija polimerazom

Kako bi umnožila egzon 2 DRB lokusa MHC sustava klase II, provela sam PCR. Ciljani odsječak DNA dugačak je oko 300 pb. Za amplifikaciju sam koristila komercijalni paket „GoTaq® Master Mix“ (Promega), a volumen reakcijske smjese iznosio je 20 μ L. Komercijalni paket sadržavao je GoTaq® Master Mix (PCR pufer s dodatkom $MgCl_2$, dNTP i HotStar Taq DNA polimeraza) i vodu bez nukleaze (Nuclease-Free Water). U konačnoj reakcijskoj smjesi GoTaq® Master Mix bio je 1x koncentriran, dok sam iz početnih koncentracija otopina početnica od 100 μ M pripremila razrjeđenje radne otopine koncentracije 2 μ M. U konačnoj reakcijskoj smjesi svaka je početnica bila koncentracije 0,2 μ M. Volumen uzorka DNA iznosio je 2 μ L, dok je ostatak reakcijske smjese do 20 μ L bila voda.

Početnice korištene za umnažanje ciljanog lokusa LA31:5'-GATCCTCTCTCTGCAGCACATTTCT i LA32:5'-TTCGCGTCACCTCGCCGCTG dizajnirane su za goveda te su preuzete iz literature (Sigurdardottir i sur. 1991). Temperaturni profil sastojao se od aktivacije GoTaq® DNA polimeraze na 95 °C tijekom 2 minute. Nakon aktivacije uslijedila su 34 ciklusa umnožavanja odsječka DNA. Prvi korak bila je denaturacija DNA na 94 °C 30 sekundi. Zatim slijedi vezivanje početnica (eng. annealing) na komplementarne odsječke DNA na temperaturi od 55 °C 30 sekundi. U trećem koraku temperatura raste na 72 °C pri čemu se tijekom 1 minute vrši elongacija (eng. extension) odnosno sinteza novih lanaca DNA. Završno produljenje DNA lanaca trajalo je tijekom 10 minuta na 72 °C. Nakon amplifikacije temperatura se spustila na 4 °C kako bi sačuvala uzorke do daljnje analize.

2.4. Elektorforeza u agaroznom gelu

S ciljem provjere uspješnosti PCR reakcije, provela sam elektroforezu u 1 % agaroznom gelu koji sam pripremila na način da sam na plameniku otopila 2 g agaroze u 20 mL 0,5X TBE (eng. Tris-Borate-Ethylenediaminetetraacetic acid) pufera. Nakon što se otopina malo ohladila dodala sam 2 μ L Syber Safe (Invitrogen) boje kako bih mogla vizualizirati željene odsječke DNA (PCR produkte). Otopinu sam zatim izlila u kadnicu te stavila češljic da se nakon polimerizacije, koja je trajala 20 minuta, formiraju jažice. Polimerizirani gel stavila sam, potom, u kadnicu koju sam prije napunila 0,5X TBE puferom. Prije nanošenja, 3 μ L svakog uzorka pomiješala sam s 3 μ L pufera za nanošenje uzoraka (GLB, eng. gel loading buffer) na parafilmu. U svaku jažicu stavila sam 6 μ L tako pripremljenog uzorka. Elektorofrezu sam provela pod naponom od 400 V i 100 mA u trajanju od 20 minuta. Konačno, vizualizaciju PCR produkta provela sam pod UV svjetlom.

2.5. Molekularno kloniranje

Ovaj postupak provodi se s ciljem odvajanja gametnih faza pojedinih jedinki, odnosno kako bi se moglo identificirati sve postojeće alele u pojedinoj jedinci koja ima duplicirane lokuse ili je heterozigot. Svrha jest fizički razdvojiti sve PCR produkte koji se koamplificiraju s istim parom početnica tijekom PCR reakcije. Molekularno kloniranje provela sam korištenjem paketa „pGEM®-T Easy Vector System“ (Promega). Postupak se provodi u četiri dana tijekom kojih se odvija ugradnja željenog umnoženog i pročišćenog odsječka DNA (egzon 2 DRB lokusa) u plazmid koji se, potom, unosi u bakterijsku stanicu. Uslijed reprodukcije bakterija dolazi do formiranja kolonija, pri čemu svaku koloniju čine bakterijske stanice koje nose umnoženi insert, a koji odgovara pojedinom alelu.

U prvom danu nakon PCR reakcije provela sam pročišćavanje PCR produkta te reakciju ligacije. Ligacija podrazumijeva ugrađivanje željenog DNA fragmenta u plazmid. Ugradnju omogućuje povezivanje timina dodanog na krajeve prekinutnog plazmida te adenina na PCR produktu. Adenin dodaje termostabilna DNA polimeraza tijekom umnožavanja odsječka.

Pročišćavanje sam provela pomoću komercijalnog kompleta „Wizard® SV Gel and PCR Clean-Up System“ (Promega). U prvom koraku sam u tubice, u kojima sam provela PCR reakciju, dodala jednaki volumen otopine za vezanje DNA na membranu (Membrane Binding Solution). Nakon toga u nove označene tubice od 2 mL s kolonicama (SV Microcolumn) prebacila sam čitav

volumen iz reakcijske tubice te inkubirala 1 minutu. Potom sam tubice centrifugirala na 14 000 okretaja/min 1 minutu te odlila višak tekućine iz tubice. U drugom koraku uslijedilo je pranje DNA koja se u prvom koraku vezala na membranu. U kolonicu sam otpipetirala prvo 700 μ L otopine za ispiranje (Membrane Wash Solution) koju sam zatim centrifugirala na 14 000 okretaja/min 1 minutu. Nakon odlijevanja tekućine iz tubice kolonicu sam vratila u istu tubicu te ponovno dodala 500 μ L iste otopine. U ovom koraku centrifugirala sam na 14 000 okretaja/min 5 minuta. Nakon odlijevanja viška tekućine kolonicu sam ponovno centrifugirala na istom broju okretaja s otvorenim poklopcem centrifuge 1 minutu s ciljem isparivanja viška etanola koji je sastavni dio otopine za ispiranje. Posljednji korak bio je elucija DNA s membrane. Slijed koraka je bio taj da sam kolonicu prebacila u označenu tubicu od 1,5 mL, dodala 40 μ L vode bez nukleaze (Nuclease-Free Water) te nakon inkubacije od 1 minute provela centrifugiranje na 14 000 okretaja/min u trajanju od 1 minute. U ovom koraku došlo je do prolaska vode bez nukleaze kroz membranu kolonice te ispiranja DNA odsječka s membrane na dno tubice, zbog čega kolonicu bacamo dok tubicu pohranjujemo na 4 °C ili na -20 °C do sljedećih koraka.

Nakon pročišćavanja uslijedila je faza ligacije odsječka DNA u plazmid (vektor). Najprije sam, kako bi ligacijska reakcija bila što uspješnija, utvrdila količinu pročišćenog PCR produkta mjerenjem absorbancije spektrofotometrom BioSpec-nanoMicro-volume UV-Vis Spectrophotometer (Shimadzu). S obzirom da je koncentracija DNA u svim uzorcima bila odgovarajuća (između 30-70 ng/ μ L), u sve ligacijske reakcije dodala sam 1 μ L pročišćenog PCR produkta. Također, postavila sam i pozitivnu i negativnu kontrolu ligacijske reakcije. U pozitivnu kontrolu dodala sam kontrolni insert te ona služi kako bih provjerila koliko je ligacija bila uspješna. Na nasađenim pločama s pozitivnom kontrolom očekivala sam izrastanje samo bijelih kolonija. Negativna kontrola sadržavala je samo plazmid, bez dodanog DNA odsječka, te sam kod nje očekivala izrastanje malog broja plavih kolonija uslijed zatvaranja plazmida koji ne sadrži DNA insert. Na ploči negativne kontrole nisam očekivala izrastanje bijelih kolonija. Naime, odabir klonova koji nose ugrađeni DNA insert provodi se metodom *plavo-bijele selekcije*. Za prvu razinu selekcije odgovoran je gen na plazmidu koji nosi otpornost na ampicilin. Ako u podlogu stavimo ampicilin, na njoj će izrasti samo one bakterije koje su uspješno usvojile plazmid, odnosno one koje nose gen za rezistenciju na taj antibiotik. Sljedeća razina odnosi se na provjeru ugradnje željenog odsječka DNA u plazmid izraslih bakterija. Plazmid, osim gena za otpornost na ampicilin, sadrži i gen za enzim galaktozidazu koja, kada je aktivna, razgradnjom X-gala (modificirani oblik

glukoze koji se dodaje u hranjivu podlogu) boji kolonije bakterija u plavo. Ugradnjom odsječka DNA u plazmid dolazi do insercijske inaktivacije navedenog gena. Zbog toga enzim postaje nefunkcionalan, s obzirom da je mjesto ugradnje odsječka unutar gena, ne razgrađuje supstrat te su posljedično kolonije bijele boje. S obzirom na cilj molekularnih kloniranja, s podloge sam sakupljala samo bijele kolonije, dakle one koje su imale ugrađeni odsječak DNA.

Ligacijsku reakciju postavila sam prema uputama proizvođača iz protokola. Komponente reakcije bile su: pGEM® - T plazmid, T4 DNA ligaza, pufer, PCR produkt i voda bez nukleaze. U reakcijske tubice od 0,2 mL otpipetirala sam 5 µL ligacijskog pufera, 1 µL pGEM® - T plazmida, 1 µL PCR produkta, 1 µL T4 DNA ligaze te do konačnog volumena od 10 µL vodu bez nukleaze. Pozitivna kontrola razlikovala se po tome što sam u nju umjesto PCR produkta stavila 2 µL kontrolnog inserta, dok sam u negativnu kontrolu otpipetirala samo pufer, vektor, ligazu i vodu do volumena 10 µL. Tubice sam pohranila preko noći u hladnjak na 4°C.

Drugi dan provela sam transformaciju odnosno unos plazmida u bakteriju pomoću ligacijske reakcije pripremljene prvi dan. Svaki sam uzorak nanijela na dvije ploče jer sam htjela sakupiti što više kolonija kako bih mogla identificirati sve prisutne alele kojih kod vrste obični jelen ima od 1 do 6. Hranjive podloge izrađene su od hranjivog medija za rast bakterija (LB, eng. lysogeny broth) koji se sastoji od: 1 L vode, 10 g Bacto®-trypton-a, 5 g Bacto®-yeast extract-a, 5 g NaCl, 15 g agara i NaOH do konačne pH vrijednosti od 7.0. Također, mediju je dodan ampicilin do konačne koncentracije od 100 µL/mL te je prije izlivanja autoklaviran. Pripremljene podloge premazala sam sa 20 µL X-gala koncentracije 50mg/mL (5-bromo-4-kloro-3-indolil-β-D-galaktopiranozid) i 100 µL 100 mM induktora reakcije IPTG-a (isopropil β-D-1-thiogalaktopiranozid) po ploči. Trakicu s ligacijskom reakcijom kratko sam centrifugirala, nakon čega sam otpipetirala 2 µL smjese u sterilne tubice od 1,5 mL koje sam držala cijelo vrijeme na ledu. Istovremeno sam u ledenoj kupelji otopila i bakterijske stanice koje sam čuvala na -80°C u ledištu. Bakterije nisam miješala pipetom već laganim kuckanjem prstima iz razloga što su vrlo osjetljive te sklone oštećenju uslijed naglih preturbacija kao i promjena temperature. Bakterijske stanice koja sam koristila su kompetentni soj (podložne transformaciji) *Escherichia coli*, JM109, genotipa endA1, recA1, gyrA96, thi, hsdR17 (rk⁻, mk⁺), relA1, supE44, Δ(lac-proAB), [F' traD36, proAB, lacI qZAM15]. Nakon što su se bakterije otopile u svaku sam tubicu s ligacijskom reakcijom otpipetirala 50 µL. Unos plazmida u bakterije provela sam na način da sam ih držala na ledu 20 minuta, zatim sam ih stavila u vodenu kupelj 45-50 na 42 °C te ih ponovno vratila na led 2

minute. Nakon toga u svaku sam tubicu dodala 900 μ L tekućeg LB medija (bez agara) te ih inkubirala 1,5 h na 37 °C uz miješanje na 150 rpm. Nakon postupka transformacije bakterije sam nasadila na ploče (100 μ L uzorka na jednu ploču; svaki uzorak na dvije ploče) te ih inkubirala na 37 °C 16-24 sata.

Treći dan provela sam presađivanje bijelih kolonija u LB tekući medij (bez agara) s ciljem da dodatno umnožim bakterijske kolonije sa željenim DNA odsječkom. Kolonije sam sakupljala u epruvete koje su sadržavale 5 mL hranjivog medija te sam za sve uzorke, osim za uzorak J9GK i J2GK, sakupljala 20 kolonija. Za uzorak J9GK prikupila sam 34 različita i odvojena bakterijska klona, dok sam za uzorak J2GK sakupila 30 bakterijskih klonova. U svaku sam epruvetu nasadila samo jednu koloniju s ploče. Nakon sakupljanja epruvete sam inkubirala uz miješanje preko noći na 37 °C.

Četvrti dan sam provela izolaciju plazmida iz bakterija te njegovu purifikaciju. Postupak se sastojao od tri koraka: lize bakterijskih stanica, pročišćavanja plazmida te ispiranja plazmida pomoću komercijalnog kompleta „PureYield™ Plasmid Miniprep System“. Prvi korak sam provela tako da sam bakterijske kulture centrifugirala na 2000 okretaja/min u trajanju od 2 minute kako bi istaložila bakterijske stanice. Nakon odlijevanja nadtaloga u svaku sam epruvetu dodala 600 μ L vode te resuspendirala bez stvaranja pjene i mjehurića. Sadržaj bakterija potom sam otpipetirala u tubicu od 1,5 mL. Potom sam u svaku tubicu dodala 100 μ L otopine za lizu stanica (Cell Lysis Buffer) i promiješala okretanjem tubice 6 puta. U vremenu od 2 minute u iste tubice otpipetirala sam i 350 μ L hladne (4-6 °C) otopine za neutralizaciju (Neutralisation Solution). Tubice sam zatim centrifugirala 3 minute na maksimalnoj brzini. Nadtalog (oko 900 μ L) otpipetirala sam na kolonicu (*PureYield Microcolumn*) koju sam stavila u označenu tubicu (*Collection Tube*) te centrifugirala 15 sekundi. Tekućinu zaostalu u tubici sam odbacila, a kolonicu vratila u istu tubicu te je uslijedio korak pročišćavanja plazmida. U kolonice sam najprije otpipetirala 200 μ L otopine Endotoxin Removal Wash (ERB) te ih centrifugirala na maksimalnoj brzini 15 sekundi. Potom sam dodala 400 μ L otopine za ispiranje kolone (Column Wash Solution) i ponovno centrifugirala na maksimalnoj brzini 30 sekundi. Po završetku centrifugiranja kolonice sam prebacila u tubice od 1,5 mL te provela ispiranje plazmida dodatkom 30 mL otopine vode bez nukleaze (nuclease-free water). Nakon inkubacije od 1 minute, kolonice sam centrifugirala 15 sekundi kako bi se plazmid eluirao u tubicu. Plazmid sam čuvala u ledištu na -20 °C.

2.6. Sekvenciranje

Jedna od metoda identifikacije egzona 2 DRB alela bila je metoda Sangerovog sekvenciranja koju sam provela u MacroGen servisu (www.macrogen.com) u podružnici Amsterdam. Na sekvenciranje sam poslala uzorke DNA dobivene molekularnim kloniranjem. Prije sekvenciranja uzorci su pročišćeni. Kako bi mogla formirati konzenzusni slijed i preciznije odrediti alele, sve sam uzorke sekvencirala iz oba smjera. Osim na Sangerovo sekvenciranje, uzorke DNA poslala sam i u servise u kojima su provedene dvije metode sekvenciranja nove generacije: sekvenciranje zasnovano na sintezi (Illumina) te sekvenciranje zasnovano na promjeni pH (Ion Torrent). Podružnica tvrtke Novogene iz Ujedinjenog Kraljevstva (U. K. Subsidiary & UK Sequencing Center) provela je sekvenciranje zasnovano na sintezi na uređaju Novaseq 6000 System (Illumina), dok je Univerza na Primorskem u Kopru provelo sekvenciranje zasnovano na promjeni pH na uređaju Ion Torrent PGM system (Applied Biosystems). Sekvenciranje zasnovano na sintezi provedeno je iz oba smjera (eng. „*Paired-End DNA Sequencing*“), dok je sekvenciranje zasnovano na promjeni pH provedeno samo iz jednog smjera. Za Sangerovo sekvenciranje korištene su iste početnice kao i za lančane reakcije polimerazom (LA31 i LA32). Početnice korištene za sekvenciranje zasnovano na promjeni pH (Ion Torrent) prikazane su na Slici 5, dok je izgled početnica (metoda izvedena servisno) korištenih za sekvenciranje zasnovano na sintezi (Illumina) prikazan na Slici 6. Uzvodne početnice na obje slike obojane su žutom bojom, dok su nizvodne početnice obojane ljubičastom bojom. Uzvodna početnica za sekvenciranje zasnovano na promjeni pH (Ion Torrent) na 5' kraju sadrži adapter (uloga kod vezanja početnice za sekvenciranje), barkod i vezni odsječak, na koji se nastavlja slijed LA31 početnice, dok je nizvodna početnica sastavljena samo od adaptera (imobilizacija na mikroskopske kuglice tijekom emulzijskog PCR-a) na koji se nastavlja slijed LA32 početnice. Uzvodna početnica korištena za sekvenciranje zasnovano na sintezi (Illumina) na 5' kraju sadrži adapter P5, nakon čega slijede motivi indeksa 2 te sekvenca Rd1SP. Naposljetku, iza Rd1SP, dolaze nukleotidi početnice LA31. Između motiva Rd1SP te uzvodne početnice mogu se dodati i 4 nasumične baze koje tijekom sekvenciranja povećavaju raznolikost između nakupina. Nizvodna početnica započinje s adapterom P7 nakon čega slijedi indeks 1, sekvenca Rd2SP te na kraju nukleotidi početnice LA32. Indeksi imaju istu funkciju kao i barkodovi kod tehnologije Ion Torrent, a to je identifikacija pojedinih uzoraka koji se sekvenciraju zajedno. Uloga adaptera vezana je uz imobilizaciju na

protočno polje čipa, dok su sekvence Rd1SP te Rd2SP mjesta vezanja početnica za sekvenciranje za tehnologiju Illumina.

5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGAGCCTATTGATGATCCTCTCTCTGCAGCACATTTCCT-
5'-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATTCGCGTCACCTCGCCGCTG-

Slika 5. Shematski prikaz početnica s dodanim nukleotidnim motivima korištenih za sekvenciranje zasnovano na promjeni pH (Ion Torrent): F početnica obojana je žuto, a R ljubičasto. Plavo su označeni adapteri, crveno barkod, zeleno vezni odsječak, a crno početnica LA31, odnosno LA32. Preuzeto iz Šurina, M (2021).

5'-P5-index2-Rd1SP-NNNN-[your-primer-1]-3'
5'-P7-index1-Rd2SP-[your-primer-2]-3'

5'-AATGATACGGCGACCAACGAGATCTACAC-XXXXXXXX-ACACTCTTCCCTACACGACGCTCTTCCGATCT-NNNN-[your-primer-1]-3'
5'-CAAGCAGAAGACGGCATAACGAGAT-XXXXXXXX-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-[your-primer-2]-3'

Slika 6. Shematski prikaz početnica s dodanim nukleotidnim motivima korištenih za sekvenciranje zasnovano na sintezi (Illumina): uzvodna početnica nalazi se u žutom okviru, a nizvodna u ljubičastom. Na donjem dijelu slike prikazan je slijed nukleotida koji odgovara adapterima P5 i P7 te sekvencama Rd1SP te Rd2SP. Nukleotidi indeksa 1 i 2 označeni su slovima „X“, dok je slovima „N“ označeno mjesto dodatka nasumičnih baza unutar početnice.

Pruzeto: https://healthsci.mcmaster.ca/docs/librariesprovider23/ngs-files/amplicon_libraries_171031.pdf?sfvrsn=958c594f_2

2.9. Računalna analiza nukleotidnih sljedova

2.9.1. Analiza rezultata Sangerovog sekvenciranja u programima Bioedit i Seqscape

Programski paket SeqScape (Applied Biosystems) koristila sam za analizu nukleotidnih sljedova i identifikaciju alela. Nakon učitavanja rezultata dobivenih Sangerovim sekvenciranjem pregledala sam sekvence te ih skratila na duljine nizova od 249 pb. Elektroferograme sam, zatim, usporedila s poznatim alelima u knjižnici kako bih identificirala sve alelne faze pojedinog uzorka. Knjižnicu sekvenci poznatih alela različitih vrsta porodice jelena preuzela sam iz GenBank baze podataka (<https://www.ncbi.nlm.nih.gov/genbank/>) pri čemu sam sve sekvence sravnala te skratila na 249 pb koliko iznosi i duljina istraživanog DNA odsječka. Aleli pod nazivima Ceel-DRB*HR01 do Ceel-DRB*HR23 (pristupni brojevi MN601742.1 - MN601764.1) otkriveni su 2020. godine istraživanjem Bujanić i suradnika na teritoriju Republike Hrvatske (Bujanić i sur. 2020). Pomoću knjižnice provjeravala sam i polimorfne nukleotidne sljedove. Osim poznatih alela, identificirala sam i nove alele. U analizi sam koristila i program Bioedit (Hall, 1999) koji mi je prvenstveno služio za poravnanje (opcija ClustaIW višestruko poravnavanje), grubu identifikaciju te usporedbu sekvenci.

2.9.2. Analiza rezultata NGS sekvenciranja korištenjem bioinformatičkog alata AmpliSAT

Nukleotidne sljedove dobivene metodama NGS, metodom zasnovanoj na sintezi (Illumina), te metodom zasnovanoj na promjeni pH (Ion Torrent), obradila sam u računalnom paketu AmpliSAT (<http://evobiolab.biol.amu.edu.pl/amplisat/>). Rezultati za analizu moraju biti spremljeni u formatu FASTQ ili FASTA koji se direktno učitava u odgovarajući program paketa na internetskom pregledniku. Detalji analize u računalnom paketu AmpliSAT prikazani su u Prilogu 1.

3. REZULTATI

3.1. Sangerovo sekvenciranje i molekularno kloniranje

Metodom Sangerovog sekvenciranja i molekularnog kloniranja pronašla sam ukupno 18 alela, od čega je 10 poznato od ranije (Tablica 1). Oznakama od Ceel21-01 do Ceel 21-11 („radni“ nazivi alela) označila sam sve alele koji su u ovom istraživanju pronađeni po prvi puta, bez obzira na korištenu metodu. Aleli koji su odgovarali onima nađenim u istraživanju Bujanić i suradnika iz 2020. godine zadržali su nazive iz tog rada. Sivom pozadinom u Tablici 2. označila sam alele koji nisu nađeni metodama molekularnog kloniranja/Sangerovog sekvenciranja, ali su nađeni s nekom drugom od preostale dvije metode genotipizacije.

Za uzorak J9GK sakupila sam 34 različita i odvojena bakterijska klona, a za uzorak J2GK 30 bakterijskih klonova. U preostalih 8 uzoraka prikupila sam po 20 različitih bakterijskih klonova. Svaka kolonija sastojala se od bakterijskih stanica koje su sadržavale umnoženi insert koji odgovara pojedinom alelu. Broj pronađenih alela po jedinci kretao se između 2 i 5, dok se broj sekvenciranih klonova u kojima je pronađen pojedini alel po jedinci kretao između 1 do najviše 24. Ostatak klonova nije dao rezultate zbog pogrešaka tijekom izvođenja metoda molekularnog kloniranja ili Sangerovog sekvenciranja. Tako je u 34 sekvencirana klona uzorka DJ9GK u njih 24 pronađen alel Ceel-DRB*HR06, dok je u samo jednom klonu tog uzorka pronađen novi alel Ceel21-07 (Tablica 1).

S obzirom na sve sekvencirane bakterijske klonove svih uzoraka, najzastupljeniji su bili aleli Ceel-DRB*HR04 (18 bakterijskih klonova) i Ceel-DRB*HR12 (21 bakterijski klon), pri čemu sam svaki od alela pronašla u četiri različita uzorka. Najzastupljeniji novopronađeni alel bio je alel Ceel21-02 kojeg sam također pronašla u 4 različita uzorka (ukupno 10 bakterijskih klonova). Najmanje zastupljeni aleli u promatranim uzorcima bili su alel Ceel-DRB*HR17 iz uzorka J20B, alel Ceel21-07 iz uzorka DJ9GK te alel Ceel21-09 iz uzorka DJ24B koje sam, kao što je vidljivo u Tablici 1., pronašla samo u po jednom bakterijskom klonu.

Tablica 1. Rezultati Sangerovog sekvenciranja i molekularnog kloniranja. Unutar svakog stupca, ispod naziva uzorka, brojevima su prikazani redni brojevi klonova u kojima su pronađeni pojedini aleli. Sivom pozadinom su označeni aleli koji nisu pronađeni ovom metodom, ali su pronađeni drugim metodama iz ovog istraživanja. Oznakama od Ceel-DRB*HR02 do Ceel-DRB*HR21 označeni su aleli koji odgovaraju alelima nađenim u istraživanju Bujanić i suradnika iz 2020. godine (Bujanić i sur. 2020), dok su oznakama Ceel21-01 do Ceel 21-11 („radni nazivi alela“) označeni aleli nađeni prvi puta u ovom istraživanju, bez obzira na korištenu metodu.

Pronađeni aleli	Jedinke				
	DJ2GK	DJ9GK	DJ16B	DJ20B	DJ24B
Ceel-DRB*HR02	1, 12, 15, 16, 19, 20, 22, 24, 27, 29			2, 6, 22	
Ceel-DRB*HR04			2, 6, 7, 10, 14, 18	4, 5, 11, 13, 21	13,16, 21, 22, 26
Ceel-DRB*HR06		1, 2, 5, 6, 8-18, 22, 24-30, 33			
Ceel-DRB*HR09					
Ceel-DRB*HR10	10, 11, 21, 28				
Ceel-DRB*HR11					
Ceel-DRB*HR12			3, 8, 11, 16	7, 8, 15, 20	7, 8, 9, 12, 15, 20, 23,24
Ceel-DRB*HR16					
Ceel-DRB*HR17				1	
Ceel-DRB*HR21					
Ceel21-01	14, 18, 23, 25, 26, 30				
Ceel21-02			9, 12, 20	14, 17	14
Ceel21-03					
Ceel21-04					
Ceel21-05					
Ceel21-06					
Ceel21-07		34			
Ceel21-08					10,11
Ceel21-09					18
Ceel21-10					
Ceel21-11					
Broj alela po jedinci	3	2	3	5	5

Nastavak Tablice 1.

Pronađeni aleli	Jedinke				
	DJ25B	DJ29B	DJ30B	D24	D28
Ceel-DRB*HR02	1-13,15-19				
Ceel-DRB*HR04			10, 16		
Ceel-DRB*HR06					
Ceel-DRB*HR09			1, 13, 15, 17, 18, 20, 7		
Ceel-DRB*HR10					
Ceel-DRB*HR11	14, 20			7,14	
Ceel-DRB*HR12			12, 3, 6, 8, 9		
Ceel-DRB*HR16		12, 14, 15, 16,19, 2, 20, 3, 4, 5, 7			4,12,20
Ceel-DRB*HR17					
Ceel-DRB*HR21		18, 21, 22, 9		1, 6, 8, 11, 13, 15, 17, 19	5, 14, 15, 17
Ceel21-01					
Ceel21-02			4, 5		
Ceel21-03		11, 13, 8		2, 3, 4, 5, 9, 10, 12, 16, 18, 20	
Ceel21-04					
Ceel21-05					1,2,3,8,9,11,18
Ceel21-06					7,13,19
Ceel21-07					
Ceel21-08					
Ceel21-09					
Ceel21-10					
Ceel21-11					
Broj alela po jedinci	2	3	4	3	4

3.2. Metoda NGS-a zasnovana na sintezi (Illumina)

Obradom rezultata dobivenih metodom sekvenciranja zasnovanoj na sintezi (Illumina), identificirala sam ukupno 19 alela, od čega je 9 novopronađenih. Sivom pozadinom u Tablici 2. označila sam alele koji nisu potvrđeni ovom metodom genotipizacije, ali jesu nekom od preostale dvije. Najviše neobrađenih očitavanja (Total_reads) (151 216) dobiveno je za uzorak D16B, a najmanje za uzorak DJ29B (94 414). Postotak kvalitete *Minimum Phred quality score* (Q30) kretao se između 98,27 % i 99,33 %. Računalnom analizom u programu AmpliCLEAN uklonjene su sve sekevence neodgovarajuće duljine i slabije kvalitete, zadržane su 1 227 184 sekvence od ukupno 1 298 414 sekvenci koje su dobivene sparivanjem očitavanja s oba kraja u programu AmpliMERGE. Od ukupnog broja za svaki je uzorak obrađeno prvih kvalitetnih 5000 sekvenci (*depth amplicon*). Uzorak s najvećom konačnom dubinom očitavanja alela (nakon klasteriranja i filtriranja) je J9GK (4178), te sam kod njega identificirala samo dva alela. To je ujedno i uzorak s najmanje pronđenih alela. Najviše alela pronđeno je u uzroku DJ24B, čak šest, pri čemu je najviše zastupljen alel Ceel-DRB*HR12 s dubinom od 907 očitavanja po ampliconu. Također, u uzorku DJ24B potvrdila sam dva nova do sada nepoznata alela. Najviše novih alela (tri) pronašla sam u uzorcima DJ16B, DJ29B I D28, dok u uzorku DJ25B nisam pronašla nijedan novi alel. (Tablica 2).

Tablica 2. Rezultati metode sekvenciranja zasnovane na sintezi (Illumina). Unutar svakog stupca, ispod naziva uzorka, prikazana je dubina očitavanja alela u određenom uzorku. Sivom pozadinom su označeni aleli koji nisu pronađeni ovom metodom, ali su pronađeni drugim metodama iz ovog istraživanja. Oznakama od Ceel-DRB*HR02 do Ceel-DRB*HR21 označeni su aleli koji odgovaraju alelima nađenim u istraživanju Bujanić i suradnika iz 2020. godine (Bujanić i sur. 2020), dok su oznakama Ceel21-01 do Ceel 21-11 („radni nazivi alela“) označeni aleli nađeni prvi puta u ovom istraživanju, bez obzira na korištenu metodu.

Pronađeni aleli	Jedinke									
	DJ2GK	DJ9GK	DJ16B	DJ20B	DJ24B	DJ25B	DJ29B	DJ30B	D24	D28
Ceel-DRB*HR02	929			626		1525				
Ceel-DRB*HR04			1215	764	668			1316		
Ceel-DRB*HR06		3350								
Ceel-DRB*HR09								759		
Ceel-DRB*HR10	1160									
Ceel-DRB*HR11						970			761	
Ceel-DRB*HR12			1450	993	907			765		
Ceel-DRB*HR16					537		1014			587
Ceel-DRB*HR17				393		1348				
Ceel-DRB*HR21					398		997		1753	598
Ceel21-01	1374									
Ceel21-02			599	330	273			463		
Ceel21-03					416		1073		1653	740
Ceel21-04								417		
Ceel21-05										854
Ceel21-06										698
Ceel21-07		828								
Ceel21-08										
Ceel21-09										
Ceel21-10			211				248			
Ceel21-11			100				155			
Dubina amplikona (Depth_amplicon)	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
Dubina alela (Depth_alleles)	3463	4178	3575	1697	3199	3843	3487	3720	4167	3477
Broj alela po jedinci	3	2	5	5	6	3	5	5	3	5

3.3. Metoda NGS-a zasnovana na promjeni pH (Ion Torrent)

Metodom sekvenciranja zasnovanoj na promjeni pH (Ion Torrent) pronašla sam najmanje alela, ukupno njih 11, od čega ih je 9 poznato od ranije. Naime, ovom sam metodom pronašla samo dva do sada nepoznata alela, alel Ceel21-03 te alel Ceel21-05. Sivom pozadinom u Tablici 3. označila sam alele koji nisu pronađeni ovom metodom genotipizacije, ali su pronađeni barem jednom od preostale dvije metode korištene u ovom istraživanju. Obradom rezultata u računalnom programu AmpliCLEAN uklonjene su sve sekvence niske kvalitete i neodgovarajuće duljine. Od ukupnog broja zadržanih sekvenci za svaki je uzorak obrađeno prvih kvalitetnih 5000 sekvenci (*depth amplicon*). Uzorak s najvećom dubinom očitavanja alela je DJ25B (4303) te sam kod njega identificirala tri već poznata alela (Tablica 3). Najviše alela, po četiri, pronašla sam u uzorcima DJ24B i D28, pri čemu je u uzorku DJ24B najzastupljeniji alel Ceel-DRB*HR04, a u uzorku D28 alel Ceel-DRB*HR16. Alel s najvećom dubinom od 3905 sekvenci je alel Ceel-DRB*HR06 iz uzorka DJ9GK, te je to jedini pronađeni alel u tom uzorku. Aleli pronađeni u najviše uzoraka (četiri) su aleli Ceel-DRB*HR04 i novopronađeni alel Ceel21-03.

Uzorci u kojima sam pronašla samo po jedan alel su DJ9GK i DJ30B. Također, u jednom uzorku i u vrlo maloj dubini (159 sekvenci) pronađen je i do sada nepoznat alel Ceel21-05.

Tablica 3. Rezultati metode sekvenciranja zasnovane na promjeni pH. Unutar svakog stupca, ispod naziva uzorka, prikazana je dubina očitavanja alela određenog uzorka. Sivom pozadinom su označeni aleli koji nisu pronađeni ovom metodom, ali su pronađeni drugim metodama iz ovog istraživanja. Oznakama od Ceel-DRB*HR02 do Ceel-DRB*HR21 označeni su aleli koji odgovaraju alelima nađenim u istraživanju Bujanić i suradnika iz 2020. godine (Bujanić i sur. 2020), dok su oznakama Ceel21-01 do Ceel 21-11 („radni nazivi alela“) označeni aleli nađeni prvi puta u ovom istraživanju, bez obzira na korištenu metodu.

Pronađeni aleli	Jedinke									
	DJ2GK	DJ9GK	DJ16B	DJ20B	DJ24B	DJ25B	DJ29B	DJ30B	D24	D28
Ceel-DRB*HR02	2896			1716		3287				
Ceel-DRB*HR04			3598	1678	2052			3903		
Ceel-DRB*HR06		3905								
Ceel-DRB*HR09										
Ceel-DRB*HR10	1115									
Ceel-DRB*HR11						608			1547	
Ceel-DRB*HR12			238	641	1035					
Ceel-DRB*HR16					548		1728			1375
Ceel-DRB*HR17						408				
Ceel-DRB*HR21							188		275	265
Ceel21-01										
Ceel21-02										
Ceel21-03					162		701		918	770
Ceel21-04										
Ceel21-05										159
Ceel21-06										
Ceel21-07										
Ceel21-08										
Ceel21-09										
Ceel21-10										
Ceel21-11										
Dubina amplikona (Depth_amplicon)	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
Dubina alela (Depth_alleles)	4011	3905	3836	4035	3797	4303	2617	3903	2740	2569
Broj alela po jedinci	2	1	2	3	4	3	3	1	3	4

3.4. Usporedba rezultata dobivenih različitim metodama genotipizacije

Alele dobivene metodom Sangerovog sekvenciranja i molekularnog kloniranja, metodom NGS zasnovanoj na sintezi (Illumina) te metodom NGS zasnovanoj na promjeni pH (Ion Torrent) prikazala sam usporedno u Tablici 4. Ukupni broj nađenih alela u svih 10 uzoraka, neovisno o upotrebljenoj metodi, iznosio je 44. Od toga je samo 50 % njih pronađeno sa sve tri metode genotipizacije. Samo metodom kloniranja/Sangerovog sekvenciranja i metodom NGS zasnovanoj na sintezi (Illumina) pronašla sam 10 alela (22,73 %), a čak 6 alela (13,64 %) pronašla sam samo metodom NGS zasnovanoj na sintezi (Illumina). Svega 2 alela nađena kloniranjem/Sangerovim sekvenciranjem (4,55 %) nisu potvrđena ni s jednom drugom metodom genotipizacije. Konačno, 4 alela (9,1 %), koja su nađena pomoću obje NGS metode, nisu potvrđena kloniranjem. Svi aleli koji su nađeni metodom NGS zasnovanoj na promjeni pH (Ion Torrent) potvrđeni su jednom ili objema drugim metodama.

Iz tih rezultata proizlazi da je najviše alela nađeno metodom NGS zasnovanoj na sintezi (Illumina) – čak 42 od 44 ukupna alela, što čini 95,45 %. Kloniranjem je nađeno ukupno 34 alela, što čini 77,27 %. Najmanje alela je nađeno metodom NGS zasnovanoj na promjeni pH (Ion Torrent) – samo 26, odn. 59,1 %.

Ako usporedimo po dvije metode genotipizacije međusobno, najviše poklapanja u identifikaciji alela nalazimo između kloniranja/Sangerovog sekvenciranja i metode Illumina – čak 72,73 % (od ukupno 44 alela, 32 su nađena tim dvjema metodama). Slijedi poklapanje između metoda Illumina i Ion Torrent, koje iznosi 59,1 % (26 alela je nađeno tim dvjema metodama), dok najmanje poklapanja nalazimo između broja alela nađenih kloniranjem/Sangerovim sekvenciranjem i metodom Ion Torrent, svega 50 % (22 alela), a to su ujedno aleli koji su nađeni svim trima metodama genotipizacije, odnosno svi aleli koji su nađeni ovim dvjema metodama, ujedno su nađeni i metodom Illumina.

Većina novih alela nije nađena svim trima metodama (samo 3 su nađena svim trima metodama), 7 ih je nađeno kloniranjem/Sangerovim sekvenciranjem i metodom Illumina, jedan je nađen NGS metodama Illumina i Ion Torrent, dok su 5 nađena samo metodom Illumina i dva samo kloniranjem/Sangerovim sekvenciranjem.

Ukupni broj alela po jedinci, nađen bilo kojom od metoda, iznosio je između dva (uzorak DJ9GK) i osam (uzorak DJ24B). Većina jedinstvenih alela nađena je u samo po jednoj jedinci (njih

čak 10 od ukupno 21 nađenih jedinstvenih alela), pa je tako i većina novih alela (Ceel21-01 – Ceel21-11) nađena u samo po jednoj jedinci (7 od 11 novih alela), dok su 2 alela (Ceel21-02 i -03) nađena u po čak 4 jedinke.

Tablica 4. Usporedba metode Sangerovog sekvenciranja i molekularnog kloniranja (*K.*), metode sekvenciranja zasnovane na sintezi (Illumina) (*Ill.*) te metode sekvenciranja zasnovane na promjeni pH (Ion Torrent) (*IT.*). Oznakama od Ceel-DRB*HR02 do Ceel-DRB*HR21 označeni su aleli koji odgovaraju alelima nađenim u istraživanju Bujanić i suradnika iz 2020. godine (Bujanić i sur. 2020), dok su oznakama Ceel21-01 do Ceel 21-11 („radni nazivi alela“) označeni aleli nađeni prvi puta u ovom istraživanju, bez obzira na korištenu metodu. Tablica 4. te Nastavak (legenda) Tablice 4. nalaze se na dvije sljedeće stranice.

	Jedinke																																a	b			
	D12GK		D19GK		D116B		D120B		D124B		D125B		D129B		D130B		D24		D28																		
	K.	III.	I/T.	K.	III.	I/T.	K.	III.	I/T.	K.	III.	I/T.	K.	III.	I/T.	K.	III.	I/T.	K.	III.	I/T.																
	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X															
Pronađeni aleli	X	X	X										X	X	X																			3			
Ceel-DRB*HR02													X	X	X																				4		
Ceel-DRB*HR04													X	X	X	X	X																		1		
Ceel-DRB*HR06									X	X	X																									1	
Ceel-DRB*HR09																																				1	
Ceel-DRB*HR10	X	X	X																																	1	
Ceel-DRB*HR11																																				2	
Ceel-DRB*HR12																																				4	
Ceel-DRB*HR16																																				3	
Ceel-DRB*HR17																																				2	
Ceel-DRB*HR21																																				4	
Ceel21-01	X	X																																		1	
Ceel21-02																																				4	
Ceel21-03																																				4	
Ceel21-04																																				1	
Ceel21-05																																				1	
Ceel21-06																																				1	
Ceel21-07									X	X																										1	
Ceel21-08																																				1	
Ceel21-09																																				1	
Ceel21-10																																				2	
Ceel21-11																																				2	
Ukupni broj alela/jedinci	3				2				5				5				8				3			5									5		44	100	
Broj alela nađen sa sve 3 metode	2				1				3				3				2				2			3									3		22	50	
Broj alela nađen s Klon. +Illum.	1				1				2				2				1				0			0									1		10	22,73	
Broj alela nađen s Illum.+IonT.	0				0				0				0				2				1			0									1		4	9,1	
Broj alela nađen s Klon.	0				0				0				0				2				0			0									0		2	4,55	
Broj alela nađen s Illum.	0				0				0				0				1				0			2									0		6	13,64	

Nastavak (legenda) Tablice 4.

"X"	oznaka metode kojom je identificiran pojedini alel
"a"	broj jedinki u kojima je alel nađen, te ukupni broj alela nađen pojedinom metodom, odnosno kombinacijom metoda
"b"	označen % nađenih alela od ukupnog broja alela (44)
	aleli nađeni sa sve tri metode genotipizacije
	alelni nađeni sa metodama kloniranja/Sangerovog sekvenciranja te metodom NGS zasnovanoj na sintezi (Illumina)
	aleli nađeni metodama NGS zasnovanih na sintezi (Illumina) te promjeni pH (Ion Torrent)
	aleli nađeni samo s metodama molekularnog kloniranja/Sangerovog sekvenciranja
	aleli nađeni samo s metodom NGS zasnovanoj na sintezi (Illumina)

4. RASPRAVA

Iako je imunološka i genetička važnost gena glavnog sustava tkivne podudarnosti neosporiva, metode za takva genetska istraživanja, posebice kod divljih vrsta životinja, još uvijek se razvijaju. Metode Sangerovog sekvenciranja i molekularnog kloniranja godinama su smatrane zlatnim standardom za istraživanja MHC sustava, no sve više ih zbog nedostataka zamjenjuju metode sekvenciranja nove generacije poput sekvenciranja zasnovanog na sintezi (Illumina) te sekvenciranja zasnovanog na promjeni pH (Ion Torrent). Tijekom ovog istraživanja, kako bi ih se usporedilo, za genotipizaciju MHC sustava korištene su sve tri spomenute metode. Genotipizaciju često otežava duplikacija DRB lokusa koju susrećemo i kod vrste obični jelen *Cervus elaphus* (Linnaeus, 1758). Duplikacija, odnosno multiplikacija lokusa i izraziti polimorfizam gena MHC sustava kod vrste obični jelen prvi su puta potvrđeni istraživanjem iz 1995. godine (Swarbrick i sur. 1995). U studiji su u 50 uzoraka nesrodnih jedinki kloniranjem i Sangerovim sekvenciranjem nađene 34 različite sekvence. Po uzorku je sekvencirano 15 bakterijskih klonova te su nađena od 1 do 4 različita alela po jedinci. Ako te rezultate usporedimo s našima dobivenim kloniranjem i Sangerovim sekvenciranjem, gdje smo u 10 nesrodnih uzoraka sekvencirali po 20 bakterijskih klonova (osim u uzorcima J9GK – 34 klona i J2GK – 30 klonova), broj alela po jedinci kretao se između 2 i 5, što je za jedan alel više naspram istraživanja iz 1995. godine. Razlog tome jest vjerojatno broj sekvenciranih klonova, koji je u našem istraživanju bio veći, te taj da se u našem istraživanju radi o populaciji koja ima dodatna 2 lokusa (ukupno tri kopije lokusa DRB), a ne samo jedan dodatni, na što ukazuju rezultati Swarbrick i sur. (1995).

Studije koje su koristile preostale dvije metode genotipizacije (metodu sekvenciranja zasnovanu na sintezi te metodu sekvenciranja zasnovanu na promjeni pH) na modelu MHC sustava vrsta iz porodice jelena, također su objavljene. Buczek i sur. (2016) su metodom sekvenciranja zasnovanoj na sintezi (Illumina) u 155 ispitivanih jedinki pronašli 46 alelnih varijanti, te od 1 do 4 različita alela po jedinci. Tom smo metodom u našem istraživanju pronašli najviše alelnih varijanti, čak 19 od kojih je 9 do sada bilo nepoznato. Broj alela po jedinici kretao se između 2 i 6 što je za dva alela po jedinci više u usporedbi sa studijom Buczek i sur. (2016). Naši rezultati dobiveni ovom metodom također ukazuju da se lokus DRB, barem u nekih od istraživanih uzoraka jelena, nalazi u tri kopije u genomu.

Metodom zasnovanoj na promjeni pH (Ion Torrent) analizirana su 23 obična jelena u studiji Bujanić i sur. (2020). Broj pronađenih alela kretao se između 1 i 6, što se ne podudara s našim istraživanjem, gdje je platformom Ion Torrenta pronađeno najmanje alela po jedinci, svega između 1 i 4.

Istraživanja temeljena na usporedbi uspješnosti više metoda genotipizacije MHC sustava provedena su i kod drugih vrsta s multiplicitiranim lokusima. Tako je 2017. godine, kao dio istraživanja u doktorskom radu D. Razali, provedena genotipizacija MHC sustava skupine I kod vrste obični vrabac (*Passer domesticus* Linnaeus, 1758) metodom Sangerovog sekvenciranja, metodom sekvenciranja zasnovanoj na sintezi (Illumina) te metodom pirosekvenciranja (Roche 454 pyrosequencing). Obični vrabac, baš poput vrste obični jelen, posjeduje multiplicitirane, izrazito polimorfne lokuse. Uspješnijom metodom pokazala se metoda sekvenciranja zasnovana na sintezi (Illumina) u kombinaciji s metodom Sangerovog sekvenciranja. Metoda pirosekvenciranja imala je izrazito malu dubinu očitavanja alela (*allele depth*) te je pokazala statistički značajna odstupanja u ponovnim analizama istih uzoraka (Razali 2017). Slični rezultati dobiveni su 2018. godine za još jednu vrstu ptice te šišmiša s multiplicitiranim lokusima. Naime, genotipizacijom MHC sustava skupine II kod vrste modrovoljka (*Luscinia svecica* Linnaeus, 1758), usporedbom dviju metoda sekvenciranja nove generacije (metode sekvenciranja zasnovane na promjeni pH (Ion Torrent) te metode sekvenciranja zasnovane na sintezi (Illumina)), kao metodom od izbora pokazala se također metoda sekvenciranja zasnovana na sintezi (Illumina) (Rekdal i sur. 2018). U istraživanju iz 2016. godine metoda sekvenciranja zasnovana na promjeni pH (Ion Torrent) kod vrste šišmiša (*Myotis lucifugus*, Le Conte 1831) opet je pokazala vrlo slabu genetsku raznolikost MHC sustava, iako su istraživači kao uzrok tome naveli malu količinu sakupljenog uzorka. Studija Duke i sur. (2015) usporedbom analiza HLA sustava metodama Ion Torrenta i Illumine, također je potvrdila kako kvalitetnije podatke daje sekvenciranje zasnovano na sintezi (Illumina).

Metoda sekvenciranja zasnovana na promjeni pH (Ion Torrent) uspješnom se pokazala u identifikaciji alela u genomima u kojima MHC geni nisu duplicirani. Takav je genom srne (*Capreolus capreolus*, Linnaeus, 1758) koja ima samo jedan DRB lokus, zbog čega su rezultati genotipizacije bili izrazito jasni. U studiji je pronađeno 8 alela, od kojih su 3 novootkrivena (Šurina 2021).

Iako se metoda sekvenciranja zasnovana na sintezi (Illumina) prema osjetljivosti u otkrivanju novih alela pokazala kao metoda od odabira za genotipizaciju MHC sustava kod vrsta s multipliciranim lokusima, postoji vjerojatnost da se tijekom sekvenciranja na toj platformi u većoj mjeri javljaju artefakti poput primjerice „lažnih alela“. Navedeno možemo zaključiti ukoliko uspoređujemo rezultate Sangerovog sekvenciranja koje je „zlatni standard“ genotipizacije MHC sustava s rezultatima platforme Illumina, pri čemu je vidljivo da se platformom Illumina pronalaze određeni aleli koji nisu nađeni kloniranjem. Takvi su, primjerice, aleli Ceel21-10 i Ceel21-11, koji osim što nisu nađeni kloniranjem (Tablica 4), u rezultatima sekvenciranja zasnovanog na sintezi (Illumina) zastupljeni su u vrlo maloj dubini (Tablica 2), zbog čega je moguće da su ti aleli zapravo artefakti nastali tijekom umnažanja. Nasuprot toga, aleli nađeni isključivo kloniranjem, a ne i platformom Illumina, jesu aleli Ceel21-08 te Ceel21-09 (Tablice 1 i 4). Činjenica da su to aleli koji do sada nisu bili poznati, a pronađeni su u malom broju bakterijskih klonova (alel Ceel21-08 pronađen je u dva klona, dok je Ceel21-09 pronađen u jednom klonu) i to u samo jednom uzorku (uzorak DJ24B), dovodi u pitanje pouzdanosti rezultata dobivenih metodom kloniranja i Sangerovog sekvenciranja u ovom uzorku. Mogućnost kontaminacije isključujemo, s obzirom da su ta dva alela nađena samo kod jedne jedinice (DJ24B). Kao razlog zašto se aleli nađeni u malom broju klonova, i to samo metodom molekularnog kloniranja i Sangerovog sekvenciranja, mogu klasificirati kao artefakti jest činjenica da se tijekom umnažanja te molekularnog kloniranja u vrsta s multipliciranim lokusima često javljaju kimere te mozaične sekvence, tj. heterodupleksi (uzrokuje ih sustav za popravak DNA bakterija tijekom molekularnog kloniranja). Formiranje kimera odvija se najčešće u zadnjim ciklusima PCR reakcije, kada su reducirane koncentracije nukleotida i početnica. Također, s obzirom da se u zadnjim ciklusima i koncentracija nepotpuno produženih novosintetiziranih lanaca naglo povećava oni formiraju *heteroduplex primers*. Konkretno, takvi nepotpuno produženi lanci ponašaju se poput početnica koje se vežu na pogrešne lance kalupe na temelju djelomične homologije pri čemu se nakon ekstenzije lanca kalupa formiraju artefaktne sekvence. Nažalost, još jedna mana tih metoda jest ta što je vrlo teško otkriti jesu li se artefakti dogodili tijekom PCR reakcije ili tijekom metode molekularnog kloniranja. Stoga, kako bi se to spriječilo, nužno je podešavanje uvjeta tijekom PCR reakcije poput podešavanja odgovarajuće temperature vezivanja početnica (eng. *annealing*), korištenja svježih pripremljenih reaktanata te smanjenja broja ciklusa PCR reakcije na broj ciklusa kada je koncentracija PCR reaktanata dovoljno visoka da limitira mogućnost nastajanja heterodupleksa. Osim toga, važno je i provođenje

više istih PCR reakcija te sekvenciranje velikog broja klonova po uzorku. S obzirom da artefakti jako nalikuju istinskim alelima, njihova pojava može dovesti do velikih pogrešaka u procjenama raznolikosti populacija (Lenz i Becker 2008; Sommer i sur. 2013). Jedan od razloga zašto se aleli u genomu s multipliciranim lokusima ne umnažaju jednako uspješno jest izrazita sličnost u sekvencama između lokusa, zbog čega je teško dizajnirati početnice specifične za svaki pojedini lokus (Sommer i sur. 2013). Iako su postojala nepodudaranja, platforma Illumina te metode molekularnog kloniranja i Sangerovog sekvenciranja potvrdile su podudaranje od čak 72,73 % (od ukupno 44 alela, 32 su nađena tim dvjema metodama).

Dodatan razlog neotkrivanja alela metodama molekularnog kloniranja i Sangerovog sekvenciranja leži također i u nejednakoj zastupljenosti alela u istraživanoj populaciji, zbog čega tijekom provođenja genotipizacije, postoji mala vjerojatnost kako ćemo odabrati i sekvencirati upravo klonove koji nose te alele (Swarbrick i sur. 1995). Razlog male vjerojatnosti odabira klona sa potencijalno novim alelom može potkrijepiti i neotkrivanje alela Ceel21-10 i Ceel21-11 metodama molekularnog kloniranja i Sangerovog sekvenciranja, već samo isključivo platformom Illumina koja je pokazala veću osjetljivost prilikom otkrivanja svih alela multipliciranih lokusa.

Metoda kojom je nađeno najmanje alela je metoda sekvenciranja zasnovana na promjeni pH (Ion Torrent), čijom smo obradom rezultata identificirali samo 11 alela, od čega ih je 9 poznato od ranije. Također, to je metoda koja je pokazala najlošije poklapanje s metodama „zlatnog standarda“ molekularnim kloniranjem i Sangerovim sekvenciranjem (50 %), a najbolje s platformom Illumina (59,1 %). Ovom su metodom identificirana samo dva nova alela, Ceel21-03 i Ceel21-05 za koje možemo zaključiti da su i istinski aleli, s obzirom da su pronađeni kod istih uzoraka i platformom Illumina.

Ukupni broj alela po jedinci, pronađen bilo kojom od metoda, kretao se između dva (uzorak DJ9GK) te osam (uzorak DJ24B) (Tablica 4). Nalaz osam alela u jednoj jedinci (DJ24B) možemo smatrat nepouzdanim, jer bi on ukazivao na to da lokus DRB u te jedinke ima čak četiri kopije, što nije nađeno niti u jedne druge jedinke u ovom ili drugim istraživanjima (najveći broj kopija lokusa DRB na koji ukazuje ovo istraživanje je tri). Stoga, najviše sumnji da su „lažni“, pokazuju aleli Ceel21-08 i Ceel21-09, iz razloga koji su ranije napisani. Stoga, smatram da uzorak DJ24B ima šest alela, te time podupire ideju da uzorci analizirani u ovom diplomskom radu imaju najviše tri kopije lokusa DRB. Većina jedinstvenih alela nađena je u samo po jednoj jedinci (10 od ukupno 21 nađenih jedinstvenih alela). Većina novih alela (njih 7 od 11) nađena je u samo po jednoj jedinci

(Tablica 4), što te rezultate ne čini vrlo pouzdanima, dok nalaz istih jedinstvenih alela u različitim jedinkama (primjerice novi aleli Ceel21-02 i Ceel21-03 nađeni su u čak 4 jedinke svaki) čini te novopronađene alele više pouzdanima.

Ukoliko uspoređujemo broj pronađenih alela s najnovijim istraživanjem iz 2020. godine (Bujanić i sur. 2020), gdje se broj alela kretao između 1 i 6, možemo zaključiti kako je, s obzirom na mali broj promatranh uzoraka, alelna raznolikost vrlo velika. Ukoliko su svi novopronađeni aleli iz ovog istraživanja istinski (osim alela Ceel21-08 i Ceel21-09 kako je ranije objašnjeno), nalaz od 6 alela po jedinci implicira na postojanje čak tri MHC lokusa kod vrste običnog jelena. Ujedno, to je i najveći broj alelnih varijanti u vrsta iz porodica jelena do sada nađen. Prijašnja istraživanja (Buczek i sur. 2016; Fernández-De-Mera i sur. 2009; Pérez-Espona i sur. 2019; Swarbrick i sur. 1995) potvrdila su postojanje najviše 4 različite alelne varijante po jedinci.

Do sada nepoznati aleli, koji su nađeni samo jednom metodom i u samo jednoj jedinci (Ceel21-04, Ceel21-08 te alel Ceel21-09, Tablica 4) smatramo vrlo nepouzdanim nalazom. Za razliku od njih, alel Ceel21-07 koji je također novopronađen u ovom istraživanju, smatramo pouzdanijim nalazom s obzirom da smo ga potvrdili dvjema metodama (molekularno kloniranje/Sangerovo sekvenciranje i platforma Illumina) i u istoj jedinci (DJ9GK). No, ipak bi bilo dobro u budućim istraživanjima dodatno potvrditi postojanje tog alela, s obzirom da je nađen u samo jednom klonu metodom Sangerovog sekvenciranja te u vrlo maloj dubini platformom Illumina (828). Također, donekle je nepouzdan i nalaz alela Ceel-DRB*HR21, koji je nađen također samo u jednoj jedinci (DJ24B) te pomoću samo jedne metode (platforma Illumina). Obzirom da je taj alel već poznat iz ranijeg istraživanja (Bujanić i sur. 2020), veća je mogućnost kako se u tom slučaju ne radi o artefaktu, već o alelu koji se jednostavno u populaciji nalazi u manjoj frekvenciji.

Najveći broj jedinstvenih alela, njih 19 (Tablica 2), pronađen je metodom sekvenciranja zasnovanom na sintezi (Illumina), dok je najmanji broj jedinstvenih alela, samo 11 (Tablica 3), nađen metodom sekvenciranja zasnovanoj na promjeni pH (Ion Torrent) (Tablica 4). S obzirom na razliku od čak osam jedinstvenih alela pronađenih između metoda, možemo zaključiti kako je metoda sekvenciranja zasnovana na sintezi (Illumina) osjetljivija prilikom identifikacije alela u dupliciranom genomu.

Rezultati ovog istraživanja ukazuju na to da je metoda sekvenciranja zasnovana na sintezi (Illumina) pouzdanija metoda genotipizacije glavnog sustava tkivne podudarnosti u genomu s

dupliciranim lokusom od metode sekvenciranja zasnovane na promjeni pH (Ion Torrent). Kao pouzdanu, platformu Illumina potvrdila su i istraživanja na drugim vrstama s dupliciranim genomom (Duke i sur. 2015; Razali 2017; Rekdal i sur. 2018).

5. ZAKLJUČAK

U ovom istraživanju usporedili smo tri vrste metoda genotipizacije MHC sustava, na primjeru DRB lokusa MHC skupine II u vrste obični jelen (*Cervus elaphus* Linnaeus, 1758). Alele egzona 2 lokusa DRB u uzorcima tkiva jelena identificirali smo korištenjem triju metoda: I. metode Sangerovog sekvenciranja i molekularnog kloniranja, II. metode NGS zasnovanoj na sintezi (Illumina), te III. metode NGS zasnovanoj na promjeni pH (Ion Torrent). Vrstu jelen obični karakterizira multiplikacija DRB lokusa, što smo uspješno potvrdili u svim jedinkama identifikacijom dva ili više alela po jedinci.

Ukupno, najviši broj pouzdano identificiranih alela po jedinci, pronađen bilo kojom od metoda, bio je šest što ukazuje na to da se lokus DRB barem u nekih od istraživanih uzoraka običnog jelena nalazi u tri kopije u genomu.

Usporedba po dvije metode sekvenciranja nove generacije, s ciljem utvrđivanja najprikladnije metode za genotipizaciju u vrsta koje imaju multiplicirane lokuse, pokazala je da je najoptimalnija metoda sekvenciranja zasnovana na sintezi (Illumina) kojom je u ovom istraživanju pronađen najveći broj jedinstvenih alela (19). Uz to, rezultati dobiveni metodom sekvenciranja zasnovanom na sintezi (Illumina) najbolje se poklapaju s rezultatima dobivenim metodama Sangerovog sekvenciranja/molekularnog kloniranja, koje se smatraju „zlatnim standardom“ u ovakvim istraživanjima. Između te dvije metode potvrđeno je najbolje poklapanje u rezultatima identifikacije alela od čak 72,73 %, tj. od ukupno 44 alela, 32 su nađena tim dvjema metodama.

Najmanji broj alela potvrđen je metodom sekvenciranja zasnovanom na promjeni pH (Ion Torrent) kojom je identificirano svega 11 alela. S obzirom na vrlo mali broj identificiranih alela, kao i mali postotak poklapanja (50 %) s metodom „zlatnog standarda“ molekularnim kloniranjem/Sangerovim sekvenciranjem, možemo zaključiti kako metoda sekvenciranja zasnovana na promjeni pH nije najoptimalnija kod genotipizacije u vrsta koje posjeduju multiplicirane lokuse.

6. LITERATURA:

- Abbas A.K., Lichtmann A.H., Pillai S. (2018): Stanična i molekularna imunologija. 399–435.
- Ambardar S., Gupta R., Trakroo D., Lal R., Vakhlu J. (2016): High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian J. Microbiol.* **56**: 394–404.
- Babik W. (2010): Methods for MHC genotyping in non-model vertebrates. *Mol. Ecol. Resour.* **10**: 237–251.
- Bentley D.R., Balasubramanian S., Swerdlow H.P., Smith G.P., Milton J., Brown C.G., Hall K.P., Evers D.J., Smith A.J. (2008): Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53–59.
- Bronner I.F., Quail M.A., Turner D.J., Swerdlow H. (2013). Improved protocols for Illumina sequencing. *Curr. Protoc. Hum. Genet.* doi:10.1002/0471142905.hg1802s79.
- Buczek M., Okarma H., Demiaszkiewicz A.W., Radwan J. (2016): MHC, parasites and antler development in red deer: No support for the Hamilton & Zuk hypothesis. *J. Evol. Biol.* **29**: 617–632.
- Bujanić M., Bužan E., Galov A., Arbanasić H., Potušek S., Stipoljev S., Šprem N., Križanović K., Konjević D. (2020): Variability of the DRB locus of MHC genes class II in red deer (*Cervus elaphus*) from a mountain region of Croatia. *Vet. Arh.* **90**: 385–392.
- Burbaite L., Csányi S. (2010): Tauri elni populiacijos ir sumediojimo pokyčiai Europoje. *Acta Zool. Litu.* **20**: 179–188.
- Duke J.L., Lind C., Mackiewicz K., Ferriola D., Papazoglou A., Derbeneva O., Wallace D., Monos D.S. (2015): Towards allele-level human leucocyte antigens genotyping - assessing two next-generation sequencing platforms: Ion Torrent personal genome machine and Illumina MiSeq. *Int. J. Immunogenet.* **42**: 346–358.
- Fernández-de-Mera I.G., Vicente J., Naranjo V., Fierro Y., Garde J.J., la Fuente J. de, Gortazar C. (2009): Impact of major histocompatibility complex class II polymorphisms on Iberian red deer parasitism and life history traits. *Infect. Genet. Evol.* **9**: 1232–1239.
- Fernández-De-Mera I.G., Vicente J., Pérez De La Lastra J.M., Mangold A.J., Naranjo V., Fierro Y., La Fuente J. De, Gortázar C. (2009): Reduced major histocompatibility complex class II

- polymorphism in a hunter-managed isolated Iberian red deer population. *J. Zool.* **277**: 157–170.
- Fuselli S., Baptista R.P., Panziera A., Magi A., Guglielmi S., Tonin R., Benazzo A., Bauzer L.G., Mazzoni C.J., Bertorelle G. (2018): A new hybrid approach for MHC genotyping: high-throughput NGS and long read MinION nanopore sequencing, with application to the non-model vertebrate Alpine chamois (*Rupicapra rupicapra*). *Heredity* (Edinb). **121**: 293–303.
- Galan M., Guivier E., Caraux G., Charbonnel N., Cosson J.F. (2010): A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics*.
- Golan D., Medvedev P. (2013): Using state machines to model the Ion Torrent sequencing process and to improve read error rates. *Bioinformatics* **29**: 344–351.
- Huang K., Zhang P., Dunn D.W., Wang T., Mi R., Li B. (2019): Assigning alleles to different loci in amplifications of duplicated loci. *Mol. Ecol. Resour.* **19**: 1240–1253.
- Kchouk M., Gibrat J.F., Elloumi M. (2017): Generations of sequencing technologies: From first to next generation. *Biol. Med.* **09**.
- Kircher M., Heyn P., Kelso J. (2011): Addressing challenges in the production and analysis of Illumina sequencing data. *BMC Genomics* **12**.
- Lenz T.L., Becker S. (2008): Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci - Implications for evolutionary analysis. *Gene* **427**: 117–123.
- Lighten J., Oosterhout C. Van, Bentzen P. (2014): Critical review of NGS analyses for de novo genotyping multigene families. *Mol. Ecol.* **23**: 3957–3972.
- Lima-Junior J. da C., Pratt-Riccio L.R. (2016): Major histocompatibility complex and malaria: Focus on *Plasmodium vivax* Infection. *Front. Immunol.* **7**: 1–13.
- Liu L., Li Y., Li S., Hu N., He Y., Pong R., Lin D., Lu L., Law M. (2012): Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**.
- Lovari S., Lorenzini R., Masseti M., Pereladova O., Carden R.F., Brook S.M., Mattioli S. (2018): *Cervus elaphus* (errata version published in 2019). *IUCN Red List Threat. Species* 2018

e.T5599707: 18.

- Mardis E.R. (2013): Next-generation sequencing platforms. *Annu. Rev. Anal. Chem.* **6**: 287–303.
- Maxam A.M., Gilbert W. (1992): A new method for sequencing DNA. 1977. *Biotechnology* **24**: 99–103.
- Pérez-Espona S., Goodall-Copestake W.P., Savirina A., Bobovikova J., Molina-Rubio C., Pérez-Barbería F.J. (2019): First assessment of MHC diversity in wild Scottish red deer populations. *Eur. J. Wildl. Res.* **65**: 1–13.
- Prober J.M., Trainor G.L., Dam R.J., Hobbs F.W., Robertson C.W., Zagursky R.J., Cocuzza A.J., Jensen M.A., Baumeister K. (1987): A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* (80-.). **238**: 336–341.
- Razali D.H.P. (2017): Spatial and temporal variation in MHC class I genes in the house sparrow (*Passer domesticus*). at <<http://etheses.whiterose.ac.uk/16473/>>.
- Rekdal S.L., Anmarkrud J.A., Johnsen A., Lifjeld J.T. (2018): Genotyping strategy matters when analyzing hypervariable major histocompatibility complex-experience from a passerine bird. *Ecol. Evol.* **8**: 1680–1692.
- Reuter J.A., Spacek D. V., Snyder M.P. (2015): High-Throughput Sequencing Technologies. *Mol. Cell* **58**: 586–597.
- Rothberg J.M., Hinz W., Rearick T.M., Schultz J., Mileski W., Davey M., Leamon J.H., Johnson K., Milgrew M.J., Edwards M., Hoon J., Simons J.F., Marran D., Myers J.W., Davidson J.F., Branting A., Nobile J.R., Puc B.P., Light D., Clark T.A., Huber M., Branciforte J.T., Stoner I.B., Cawley S.E., Lyons M., Fu Y., Homer N., Sedova M., Miao X., Reed B., Sabina J., Feierstein E., Schorn M., Alanjary M., Dimalanta E., Dressman D., Kasinskas R., Sokolsky T., Fidanza J.A., Namsaraev E., McKernan K.J., Williams A., Roth G.T., Bustillo J. (2011): An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**: 348–352.
- Shiina T., Suzuki S., Kulski J.K. (2016): MHC Genotyping in Human and Nonhuman Species by PCR based Next-Generation Sequencing. *Next Gener. Seq. - Adv. Appl. Challenges* doi:10.5772/61842.

- Sigurdardottir S., Borsch C., Gustafsson K., Andersson L. (1991): Cloning and sequence analysis of 14 DRB alleles of the bovine major histocompatibility complex by using the polymerase chain reaction. *Anim. Genet.* **22**: 199–209.
- Sommer S. (2005): The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front. Zool.* **2**: 1–18.
- Sommer S., Courtiol A., Mazzoni C.J. (2013): MHC genotyping of non-model organisms using next-generation sequencing: A new methodology to deal with artefacts and allelic dropout. *BMC Genomics* **14**.
- Stipoljev S., Bužan E., Rolečková B., Iacolina L., Šprem N. (2020): MHC genotyping by SSCP and amplicon-based NGS approach in chamois. *Animals* **10**: 1–9.
- Swarbrick P.A., Crawford A.M. (1997): The red deer (*Cervus elaphus*) contains two expressed major histocompatibility complex class II DQB genes. *Anim. Genet.* **28**: 49–51.
- Swarbrick P.A., Schwaiger F.W., Epplen J.T., Buchan G.S., Griffin J.F.T., Crawford A.M. (1995): Cloning and sequencing of expressed DRB genes of the red deer (*Cervus elaphus*) Mhc. *Immunogenetics* **42**: 1–9.
- Šurina M. (2021): Genotipizacija lokusa glavnog sustava tkivne podudarnosti srne, *Capreolus capreolus* (Linnaeus, 1758) sekvenciranjem nove generacije. Diplomski rad (u postupku)
- Walker S.E., Lorsch J. (Elsevier Inc.: 2013). Sanger dideoxy sequencing of DNA. *Methods Enzymol.* **529**.

Izvori sa interneta:

1. <https://www.illumina.com/techniques/sequencing/dna-sequencing/targeted-resequencing/amplicon-sequencing.html> (Pristupljeno: 10.7.2021)
2. <https://support.illumina.com/bulletins/2020/06/illumina-adapter-portfolio.html> (Pristupljeno: 13.6.2021)
3. <https://www.youtube.com/watch?v=zBPKj0mMcDg> (Pristupljeno: 15.6.2021)
4. https://healthsci.mcmaster.ca/docs/librariesprovider23/ngsfiles/amplicon_libraries_171031.pdf?sfvrsn=958c594f_2 (Pristupljeno: 20.9.2021)
5. www.macrogen.com (Pristupljeno: 10.5.2021)
6. <http://evobiolab.biol.amu.edu.pl/amplisat/> (Pristupljeno: 4.4.2021)

7. ŽIVOTOPIS

Rođena sam 16. 4. 1996. godine u Koprivnici, dok sam djetinjstvo provela u malom mjestu u blizini Križevaca, gdje živim i danas. Do 2011. godine pohađala sam Osnovnu školu Kalnik, nakon koje iste godine upisujem srednju Medicinsku školu u Bjelovaru. U 5. razredu srednje škole primila sam nagradu fonda Boža Tvrtković za najbolju učenicu škole 2016. godine. Nakon srednje škole upisujem preddiplomski studij biologije na Odjelu za biologiju Sveučilišta J. J. Strossmayera u Osijeku. Prediplomski studij završavam 2019. godine te iste godine odlazim na diplomski studij Eksperimentalne biologije na Biološkom odsjeku Prirodoslovno-matematičkog fakulteta u Zagrebu. Privrženost prema biologiji, posebno granama fiziologije te genetike, razvijala sam brojnim izvannastavnim aktivnostima tokom cijelog školovanja te bih se u tom smjeru voljela usavršavati i u budućnosti.

8. PRILOZI

Prilog 1. Detaljan prikaz analize u računalnom paketu AmpliSAT
(<http://evobiolab.biol.amu.edu.pl/amplisat/>)

Analizu rezultata dobivenih metodom sekvenciranja sintezom (Illumina) započela sam u programskom alatu AmpliMERGE koji formira konsenzus slijed iz nukleotidnih sljedova sekvenciranih uzvodnom i nizvodnom početnicom. Rezultati dobiveni sekvenciranjem za svaki smjer nalaze se u odvojenim datotekama koje se unose u filter nakon što se komprimiraju u ZIP format (Prilog - slika 1.). Rezultate za svaki smjer potrebno je komprimirati u odvojenim datotekama, odnosno rezultate sekvenciranja uzvodnom početnicom stavljamo u jednu datoteku, dok rezultate sekvenciranja nizvodnom početnicom unosimo u drugu odvojenu komprimiranu datoteku. Unos podataka isključivo je u ZIP formatu. Dodatne parametre u ovom koraku nisam postavila.

Run AmpliMERGE

Run name:

Email (optional):

Paired-end reads file R1: FASTQ (compressed or uncompressed)
Max. 500 MB [Download example](#)
 10jelena_R1.zip

If reads have been already de-multiplexed into separate files (one file per sample), you can pack them into a single .zip or .tar.gz format file and use it as input. In such case file 1 and file 2 must have identical names and differ only in the suffixes 'R1' and 'R2'.

Paired-end reads file R2: FASTQ (compressed or uncompressed)
Max. 500 MB [Download example](#)
 10jelena_R2.zip

[Advanced program parameters](#)

Prilog - slika 1. Analiza nukleotidnih sljedova u AmpliMERGU (Illumina). Crvenim strelicama označena su polja u koja se unose datoteke s rezultatima dobivenih tehnologijom sekvenciranja sintezom (Illumina).

Datoteku dobivenu AmpliMERGE-om potom sam učitala u programski alat AmpliCLEAN. On uklanja sve sekvence koje nisu odgovarajuće duljine. Osim toga, brišu se i sve nekvalitetne sekvence prema postavljenom parametru *Minimum Phred quality score*. Kriteriji koje sam postavila bili su: *Minimum Phred quality score* 30 (Q30), minimalna duljina sekvence (*Minimum sequence length*) 307 pb te maksimalna duljina sekvence (*Maximum sequence length*) 317 pb

Prilog – slika 2.). Vrijednost Q30 označava pouzdanost očitavanja od 99.9 %, tj. mogućnost pogrešno očitane baze u omjeru 1:1000. U *Amplicon data* upisala sam:

```
>marker,primer_f,primer_r,gene,feature,species
MHC,250,GATCCTCTCTCTGCAGCACATTTCCT,TTCGCGTCACCTCGCCGCTG,DRB MHC2,exon 2,jelen
>sample,tag_f,tag_r
D24,AAACATCG,ACAAGCTA
D28,AAACATCG,CTGTAGCC
DJ2GK,AACGTGAT,CATCAAGT
DJ9GK,AACGTGAT,CTGTAGCC
DJ16B,AACGTGAT,AGTACAAG
DJ20B,AACGTGAT,ACAAGCTA
DJ24B,AACGTGAT,AACAACCA
DJ25B,AAACATCG,CGCTGATC
DJ29B,AACGTGAT,CGCTGATC
DJ30B,AAACATCG,CATCAAGT
```

Run AmpliCLEAN

Run name:

Email (optional):

Sequence file: FASTA/FASTQ (compressed or uncompressed) without spaces, parenthesis or other special symbols in the name. If reads have been already de-multiplexed into separate files (one file per sample), you can pack them into a single .zip or .tar.gz format file and use it as input. In such case you should leave empty the amplicon data section. Max. 500 MB [Download example](#)

1. 10jelenaMERGE tar.gz

Amplicon data: **2.**

It is very important to specify all the primer and tag sequences in 5'->3' sense. Shortening primer sequences to 7-9 nts can increase the number of retrieved sequences (eg. GAGTGTGAT instead of GAGTGTGATTTCTCAACGGGA).

3. Minimum Phred quality score: Reads/sequences with lower average Phred quality score will be removed. Only Sanger/Phred+33 encoding format is supported.

Minimum sequence length: Reads/sequences with lower lengths will be removed.

Maximum sequence length: Reads/sequences with higher lengths will be removed.

Trim sequence length: Reads/sequences will be trimmed to the desired length.

Maximum number of reads per amplicon: Maximum number of amplicon sequences/reads to be processed.

Prilog - slika 2. Analiza nukleotidnih sljedova u AmpliCLEANU-u. Crvenim strelicama označena su polja za unos: 1.) datoteke dobivene analizom u AmpliMERG-u, 2.) podataka o ampliconu (*Amplicon data*) te 3.) željenih parametara.

Datoteku s rezultatima iz AmpliCLEAN-a učitala sam u AmpliCHECK koji je namijenjen preliminarnoj analizi podataka sekvenciranja. Filteri provode demultipleksiranje podataka te brzu analizu visokofrekventnih varijanti u svakome klasteru. U AmpliCHECK-u zadala sam samo dva parametara: *Minimum amplicon frequency*: 1 %, *Maximum number of reads per amplicon*: 5000. Kao vrstu tehnologije odabrala sam opciju *Illumina* (Prilog – slika 3.). Na temelju podataka dobivenih AmpliCHECK-om, postavila sam parametre u daljnjoj analizi u programskom alatu AmpliSAS. Pravilno postavljanje parametara analize bitno je kako bi se odbacili svi potencijalni artefakti, a zadržale prave sekvence. Također, potrebno je spriječiti klasifikaciju slabo zastupljenih „pravih“ alela u artefakte. U AmpliCHECK učitala sam referentnu knjižnicu alela za usporedbu s dobivenim sekvencama kako bi se identificirali već poznati i potencijalno novi aleli.

Run AmpliCHECK

Run name:

Email (optional):

Sequence file: FASTA/FASTQ (compressed or uncompressed) without spaces, parenthesis or other special symbols in the name. If reads have been already de-multiplexed into separate files (one file per sample), you can pack them into a single .zip or .tar.gz format file and use it as input. In such case you should leave empty the amplicon data section. Max: 500 MB [Download example](#)

1. nology: ☐ Auto ☒ Illumina ☐ 454/IonTorrent Analysis parameters will be optimized for the selected sequencing technology, they can be modified in "Advanced options".

If transcriptomic data is being analyzed or many co-amplifying alleles are expected (>10, e.g. MHC class I), the 'minimum per amplicon frequency' threshold in 'Advanced program parameters' section should be manually adjusted.

Expand results: ☐ Yes ☒ No If selected this option, the results will be printed in 3 spreadsheets (depths, frequencies and errors).

2. Amplicon data: It is very important to specify all the primer and tag sequences in 5'->3' sense. Shortening primer sequences to 7-9 nts can increase the number of retrieved sequences (eg. GAGTGTTCAT instead of GAGTGTTCATTCTCCAACGGGA).

3. or file: Max: 100 KB [See example](#)

4. file FASTA format Max: 2 MB [See example](#)

☒ [Advanced program parameters](#)

Analysis parameters:
Leave in blank if you don't want to use the parameter.

Substitution errors (%): % Unique sequences with lower number of substitutions will be annotated as artefacts of previous sequences (substitutions = errors x length).

Indel errors (%): % Unique sequences with lower number of non-homopolymer indels will be annotated as artefacts of previous sequences (indels = errors x length).

Identity threshold (%): % Unique sequences with higher identity will be annotated as artefacts of previous sequences.

Minimum amplicon frequency (%): % Unique sequences with lower frequency will be discarded.

Minimum variant depth: Unique sequences with lower coverage will be discarded.

Minimum amplicon depth: Amplicons with lower total coverage will be discarded.

Maximum number of reads per amplicon: Maximum number of amplicon sequences/reads to be processed.

Prilog - slika 3. Analiza nukleotidnih sljedova u AmpliCHECK-u. Crvenim strelicama označena su polja za unos: 1.) datoteke dobivene analizom u AmpliCLEAN-u, 2.) podataka o ampliconu (*Amplicon data*), 3.) knjižnice alela te 4.) željenih parametara.

Zadnji korak je iscrpna analiza alela i genotipizacija prema odabranim parametrima na temelju dobivenih podataka iz AmpliCHECKA u filterima programskog alata AmpliSAS. Proces rada AmpliSAS-a slijedi tri koraka: de-multipleksiranje (*de-multiplexing*), klasteriranje (*clustering*) i filtriranje (*filtering*) jedinstvenih sekvenci u svrhu genotipizacije. De-multipleksiranje sekvence grupira u amplikone prema početnicama i oznakama (eng. tag). Nadalje, sekvence u amplikonima podvrgavaju se klasteriranju gdje se svi artefakti i potencijalni aleli kategoriziraju prema parametrima ograničenja postavljenim od strane istraživača. Parametre je potrebno postaviti i za korak filtriranja koji odbacuje ili zadržava okvir čitanja (*frameshifts*) te kimerne sekvence (*chimeras*) – amplikoni koji sadrže motive sekvence sastavljene od dijelova dvije ili više istinskih sekvenci. Također, korak filtriranja vrši analizu s obzirom na zadanu *Minimum variant depth*, *Minimum amplicon frequency* te *Maximum amplicon length deviation*. U AmpliSAS učitala sam datoteku s podacima dobivenim u alatu AmpliCLEAN. Također, postavila sam sljedeće parametre: maksimalan broj alela po amplikonu (*Maximum number of alleles per amplicon*) 10, minimalna dubina amplikona (*Minimum amplicon depth*) 100, razlog odbacivanja sekvence (*Verbose mode*) - YES. Dodatno, postavila sam maksimalan broj očitavanja po amplikonu (*Maximum number of reads per amplicon*) 5000. U koraku stvaranja nakupina (eng. clustering parameters) odabrala sam postotak pogreške supstitucijom (*Substitution errors*) 1 %, indela 0,001 %, minimalnu dominantnu frekvenciju (*Minimum dominant frequency*) 25 % i minimalnu frekvenciju amplikona (*Minimum amplicon frequency*) 1 %. Svi parametri, osim *Minimum amplicon frequency*, postavljeni su prema već zadanim parametrima za tehnologiju *Illumina*. Za parametar *Minimum amplicon frequency*, umjesto zadanih 3 %, odabrala sam parametar od 1 % kako bih pronašla i one slabo zastupljene istinske alele koji bi se, da je postavljen parametar od 3 %, svrstali u artefakte. U kriterijima filtriranja odabrala sam opciju odbaci kimere (*Discard chimeras*). Podaci o amplikonu (*Amplicon data*) i referentna knjižnica alela bili su identični kao u filteru AmpliCHECK (Prilog – slika 4.).

Run AmpliSAS

Run name: 10jelenSAS
 Email (optional): anagalov@biol.pmf.hr
 Sequence file: **1.** Choose File 10jelenACLEAN.tar.gz
FASTA/FASTQ (compressed or uncompressed) without spaces, parenthesis or other special symbols in the name. If reads have been already de-multiplexed into separate files (one file per sample), you can pack them into a single .zip or .tar.gz format file and use it as input. In such case you should leave empty the amplicon data section. Max. 500 MB Download example
 Or file URL:
 Technology: ☐ Auto ☒ Illumina ☐ 454/IonTorrent Analysis parameters will be optimized for the selected sequencing technology; they can be modified in "Advanced options".
 Maximum number of alleles per amplicon: **2.** 10 Maximum number of expected alleles per amplicon, it will limit the number of clusters to calculate and also this number will be used for Degree of Change calculations in case this filtering option is chosen.
 Minimum amplicon depth: 100 Amplicons with lower total coverage will be discarded.
 Amplicon data: **3.** It is very important to specify all the primer and tag sequences in 5'->3' sense. Shortening primer sequences to 7-9 nts can increase the number of retrieved sequences (eg. GAGTGTGCAT instead of GAGTGTGATTTTCCACCGGA).

```
>marker,length,primer_f,primer_r,gene,feature,species
HMC,250,GATCCTCTCTCTGACACATTTCCT,TTCCGTGACCTGCGCCGCTG,DRB HMC2,exon 2,jelen
>sample,tag_f,tag_r
D24,AAACATCG,ACAAAGCTA
D28,AAACATCG,CTGTAGCC
D126K,AACTGAT,ATCAAGT
D196K,AACTGAT,CTGTAGCC
D116B,AACTGAT,AGTACAAG
D120B,AACTGAT,ACAAAGCTA
D124B,AACTGAT,ACAAACCA
D125B,AAACATCG,CGCTGATC
D129B,AACTGAT,CGCTGATC
D130B,AAACATCG,CATCAAGT
```


 or file: Max. 100 KB See example
 Choose File No file chosen
 Alleles file (optional): FASTA format Max. 2 MB See example
 Choose File knjiznica-jeleni.fasta
Additional options:
 Only de-multiplexing: No ☐ If selected this option, sequences/reads will be de-multiplexed into amplicons (very fast, clustering and filtering steps will be skipped).
 Skip clustering: No ☐ If selected this option, clustering step will be skipped and clustering parameters will be ignored.
 Skip filtering: No ☐ If selected this option, filtering step will be skipped and filtering parameters will be ignored.
 Verbose mode: Yes ☒ All sequences clustered and/or filtered will be printed into individual files explaining the clustering/filtering reasons.
 Maximum number of reads per amplicon: 5000 Maximum number of amplicon sequences/reads to be processed.
[Advanced program parameters](#)
Clustering parameters:
Leave in blank if you don't want to use the parameter.
 Substitution errors (%): **4.** 1 % Sequences with higher number of substitutions will be classified into new clusters (substitutions = errors x length).
 Indel errors (%): 0.001 % Sequences with higher number of non-homopolymer indels will be classified into new clusters (indels = errors x length).
 Clustering identity threshold (%): % Sequences with lower identity will be classified into new clusters.
 Minimum dominant frequency (%): 25 % Sequences within a cluster with same or higher frequency respect to the dominant will be classified as 'subdominants' and moved to a new cluster.
 Minimum amplicon frequency (%): % Sequences with same or higher frequency within the same amplicon will be classified as 'subdominants' and moved to a new cluster.
 Exact length required: Yes ☒ Only sequences that satisfy theoretical marker lengths can be 'dominant' within a cluster.
 In-frame length required: No ☒ Only sequences in-frame with marker theoretical lengths can be 'dominant' within a cluster.
 Keep singletons: No ☒ Keep singletons (OTUs with coverage 1) after clustering.
 Cluster correction with alleles: Yes ☒ Sequences given in the allele file will be used to correct 'dominant' ones within clusters.
Filtering parameters:
Leave in blank if you don't want to apply the filter.
 Minimum variant depth: Sequences with lower coverage will be discarded.
 Minimum amplicon frequency (%): 1 % Sequences with lower frequency will be discarded.
 Maximum amplicon length deviation: Sequences longer or shorter than the marker theoretical lengthszvalue will be discarded.
 Discard chimeras: Yes ☒ Sequences that are chimeras from other major sequences will be discarded.
 Discard noncoding: No ☒ Sequences with stop codons in major reading frame will be discarded.
 Discard frameshifts: No ☒ Sequences not in-frame with marker theoretical lengths will be discarded.
 Commonness: Samples: Frequency: Sequences present in an equal or higher number of samples will be kept, if they pass the minimum specified frequency, even if they do not pass other filters.
 Degree of Change (DOC): No ☒ If DOC option is chosen, all other filters will be ignored. Remember to adjust the parameter 'Maximum number of alleles per amplicon' above.

Run

Reset

Prilog - slika 4. Analiza nukleotidnih sljedova u AmpliSAS-u. Crvenim strelicama označena su polja za unos: 1.) datoteke dobivene analizom u AmpliCLEAN-u, 2.) podataka o ampliconu (*Amplicon data*), 3.) knjižnice alela te 4.) željenih parametara

Nukleotidni sljedovi dobiveni obradom u tehnologiji Ion Torrenta također su obrađeni u programskom paketu AmpliSAT. Razlika u obradi podataka naspram Illumine temeljila se na promjeni nekoliko parametara filtera te klasteriranja. Osim toga, analiza u programu AmpliMERGE ne provodi se, a u odabiru tehnologije korištene za sekvenciranje, umjesto *Illumina*, postavila sam opciju *Ion Torrent*. Također, razlike su bile i u postavkama programa AmpliCHECK: postotak greške supstitucijom (*Substitution errors*) 0,5 %, indela 1 %, minimalne frekvencije amplikona (*Minimum amplicon frequency*) 1 %. U AmpliSAS-u odabrala sam: postotak pogreške supstitucijom (*Substitution errors*) 0,5, indela 1 %, minimalnu dominantnu frekvenciju (*Minimum dominant frequency*) 25 % i minimalnu frekvenciju amplikona (*Minimum amplicon frequency*) 3 %. Podaci o amplikonu bili su: marker,primer_f,primer_r,gene,feature,species (MHC,CCTCTCTCTGCAGCACATTTTCCT,TTCGCGTCACCTCGCCGCTG,DRB,MHC2,exon 2,Cervus elaphus), te u drugom redu: sample,tag_f,tag_r,(naziv uzorka,GAT).