UNIVERZA V LJUBLJANI

BIOTEHNIŠKA FAKULTETA

Jana OBŠTETER

**UČINKOVITA UPORABA GENOMSKIH INFORMACIJ V REJSKIH PROGRAMIH ZA MAJHNE POPULACIJE GOVEDI**

DOKTORSKA DISERTACIJA

Doktorski študij Bioznanosti

Ljubljana, 2020

UNIVERZA V LJUBLJANI

BIOTEHNIŠKA FAKULTETA

Jana OBŠTETER

**UČINKOVITA UPORABA GENOMSKIH INFORMACIJ V REJSKIH PROGRAMIH ZA MAJHNE POPULACIJE GOVEDI**

DOKTORSKA DISERTACIJA

Interdisciplinarni doktorski študijski program Bioznanosti

**EFFICIENT USE OF GENOMIC INFORMATION IN BREEDING PROGRAMS FOR SMALL CATTLE POPULATIONS**

DOCTORAL DISERTATION

Interdisciplinary doctoral study programme of Biosciences

Ljubljana, 2020

Doktorsko delo je zaključek Interdisciplinarnega doktorskega študijskega programa Bioznanosti. Delo je bilo opravljeno na Katedri za …

Senat Biotehniške fakultete je za mentorja doktorskega dela dela imenoval izr. prof. dr. Gregorja Gorjanca in za somentorja dr. Janeza Jenka*.*

Recenzent/ka:

Komisija za oceno in zagovor:

Predsednik: prof. dr. *oz. ustrezni naziv* Ime PRIIMEK

Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko

Član: prof. dr. Ime PRIIMEK

Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko

Član: prof. dr. Ime PRIIMEK

Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko

Datum zagovora:

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

ŠD Dd

DK

KG kvantitativna genetika, genomika, znanost o živalih, genomska selekcija, učinkovitost, govedo

AV OBŠTETER, Jana, Master of Science

SA GORJANC, Gregor (mentor), JENKO, Janez (somentor)

KZ SI-1000 Ljubljana, Jamnikarjeva 101

ZA Univerza v Ljubljani, Biotehniška fakulteta, Interdisciplinarni doktorski študijski program Bioznanosti

LI 2020

IN UČINKOVITA UPORABA GENOMSKIH INFORMACIJ V REJSKIH PROGRAMIH ZA MAJHNE POPULACIJE GOVEDI

TD Doktorska disertacija

OP IV, št. str., št. pregl., št. sl., št. pril., št. vir.

IJ sl/en

JI sl/en

AI *Izvleček naj vsebuje približno 200 besed (ne sme preseči te strani).*

KEY WORDS DOCUMENTATION

ND Dd

DC

CX quantitative genetics, genomics, animal breeding, selection

AU OBŠTETER, Jana, Master of Science

AA GORJANC, Gregor (supervisor), JENKO, Janez (co-advisor)

PP SI-1000 Ljubljana, Jamnikarjeva 101

PB University of Ljubljana, Biotechnical Faculty, Interdisciplinary doctoral study programme of Biosciences

PY 2020

TI EFFICIENT USE OF GENOMIC INFORMATION IN BREEDING PROGRAMMES FOR SMALL CATTLE POPULATIONS

DT Doctoral disertation

NO IV, št. p., št. tab., št. fig., št. ann., št. ref.

LA sl/en

Al sl/en

AB *Abstract (pribl. 200 besed, ne sme preseči te strani)*

KAZALO VSEBINE

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  | KLJUČNA DOKUMENTACIJSKA INFORMACIJA | II |
|  | KEY WORDS DOCUMENTATION | III |
|  | KAZALO VSEBINE |  |
|  | KAZALO PREGLEDNIC |  |
|  | KAZALO SLIK |  |
|  | KAZALO PRILOG |  |
|  | SEZNAM GESEL *opis ključnih besed, izbor naredi kandidat s pomočjo mentorja* |  |
|  | OKRAJŠAVE IN SIMBOLI |  |
| **1** | **UVOD** | 1 |
| **2** | **PREGLED OBJAV** | 2 |
| 2.1 | NASLOV PODPOGLAVJA | 3 |
| **2.1.1** | **Naslov podpoglavja** | 4 |
| 2.1.1.1 | Naslov podpoglavja |  |
| **…** |  | **…** |
| **3** | **MATERIAL IN METODE** |  |
| 3.1 | MATERIAL |  |
| 3.2 | METODE |  |
| **4** | **REZULTATI** *(ali skupno poglavje: REZULTATI Z RAZPRAVO*  *po dogovoru z mentorjem)* |  |
| **5** | **RAZPRAVA** |  |
| **6** | **SKLEPI** |  |
| **7** | **POVZETEK** |  |
| **8** | **VIRI**  **ZAHVALA** |  |
|  | **PRILOGE** |  |

# **1** UVOD S PREDSTAVITVIJO PROBLEMATIKE, CILJEV IN HIPOTEZ/ZNANSTVENIH VPRAŠANJ

## 1.1 Genetsko izboljševanje populacij

**Osnove dedovanja**

Večino lastnosti in bolezni, če ne vse, imajo genetsko komponento. Genetska informacija je kodirana na dolgih zaporedjih deoksiribonukleinske kisline (DNK), ki jih imenujemo kromosomi. Za večino zaporedja DNK še vedno ne poznamo točne funkcije. Bolje so raziskani deli, ki kodirajo proteinske produkte, t.j. geni. Zaporedje DNK ni statično, saj ga biološki procesi in evolucijske sile iz generacije v generacijo spreminjajo. Zaporedje DNK je značilno za vrsto, vendar pa najdemo razlike med posamezniki. Te se ustvarijo pri ustvarjanju nove generacije. Glavna vira genetske raznovrstnosti (variabilnosti / variacije) sta procesa mutacije in rekombinacije, ki ustvarjata razlike med posamezniki. Mutacije so posledica netočnega podvajanja molekule DNK. Za samo evolucijo so pomembne mutacije v spolnih celicah, ki se prenesejo na naslednje generacije. Druga sila v spolnih populacijah je rekombinacija. Pri procesu nastajanja spolnih celic v mejozi prihaja do rekombinacije starševskih kromosomov. V diploidnih dvospolnih populacijah dobi potomec polovico genetskega materiala od matere in polovico od očeta. Kromosomi v gamete segregirajo naključno, kar je dodaten vir genetske variabilnosti v populaciji.

Genetski ustroj organizma imenujemo njegov genotip. V populaciji nas zanimajo mesta na genomu oz. lokusi, na katerih se posamezniki med seboj razlikujejo, t.i. polimorfna mesta. Večina polimorfnih mest ima dve različici oz. alela, so bialelni, kar je v skladu s teorijo neskončnega števila mest (angl. i*nfinite site model*), ki pravi, da vsako mesto na genomu mutira natančno enkrat. Na takšnih mestih so posamezniki lahko homozigotni (imajo dve kopiji istega alela) ali heterozigotni (imajo po eno kopijo vsakega alela). Takšni lokusi se dedujejo po Mendelskih zakon, zato frekvence genotipov v populaciji sledijo Hardy‑Weinbergovem ravnovesju. Po pričakovanjih je frekvenca obeh alelov v populaciji enaka, pričakovane frekvence genotipov pa so ¼ za homozigote za referenčni alel (divji tip), ½ za heterozigote in ¼ za homozigotne za alternativni alel.

Genotip posameznika delno ali v celoti pogojuje posameznikov fenotip, ki je morfološka, fiziološka ali opažena lastnost. Nekatere izmed izraženih lastnosti so v celoti nadzorovane z genotipom in so pod vplivom enega ali majhnega števila genov. Te lastnosti so diskretne in sledijo Mendelskemu vzorcu dedovanja ter jih imenujmo Mendelske oz. kvalitativne lastnosti. Druge lastnosti so le delno nadzorovane z genotipom posameznika, delno pa z okoljskimi vplivi. Te lastnosti imajo kontinuirano porazdeljene fenotipe, ki sledijo normalni porazdelitvi, in jih imenujemo kvantitativne oz. kompleksne lastnosti. Kvantitativne lastnosti so nadzorovane z velikim številom genov, so poligenske, pri čemer nas zanima skupni učinek vseh genov. Lokusi z vplivom na fenotip imenujemo vzročni lokusi oz. kvantitativni lokusi (angl. *quantitative trait loci*), pri čemer pa lahko imajo nekateri lokusi večji (merljiv) učinek na fenotip.

Učinki alelov na fenotip so lahko additivni, posledica dominance, epistaze ali interakcije genotipa z okoljem.

## 1. 2 selekcija

Selekcija stremi h genetskemu izboljšanju populacije. Ključen korak selekcije je odbira genetsko najboljših posameznikov za starše naslednje generacije. Identifikacija teh posameznikov je težavna zaradi več virov variabilnosti. Med temi so genetska variabilnost za lastnost, velik okoljski vpliv, segregacija in rekombinacija pri potomcih že identificiranih elitnih živali ter vključenost različnih organizacij v selekcijsko delo. Uspeh selekcije merimo z doseženim genetskim napredkom (ΔG),

ΔG = *i* x *r* x *σa*/ *L*; [1]

kjer je *∆G* letni genetski napredek, merjen v enotah standardnega odklona na leto, *i* intenzivnost selekcije, *r* točnost selekcije, *σa* genetska standardna deviacija za lastnost v populaciji in *L* generacijski interval (Lush, 1945). Ti parametri se bistveno razlikujejo med moško in žensko potjo selekcije pri govedi, zato sta Rendel in Robertson, 1950, definirala štiri poti selekcije pri govedu: oče-sin, oče-hči, mati-sin in mati-hči. Najstrožja selekcija se vrši v selekciji očetov, saj je vpliv moških živali zaradi velikega števila potomcev, ki jih dosežemo preko umetnega osemenjevanja, znatno večji od vpliva ženskih živali, ki imajo v povprečju le enega potomca letno.

za izbrane lastnosti. Slednje ponavadi izberemo na podlagi njihove gospodarske pomembnosti, vendar pa morajo te lastnosti zadostiti tudi drugim kriterijem. Lastnosti, na katere odbiramo, morajo biti imeti zadostno genetsko variabilnost in imeti zadostno heritabiliteto. Morajo biti jasno definirane, imeti nizko ceno merjenja ter biti redno »recorded«. V govedoreji se je nabor lastnosti za selekcijo skozi zgodovino precej spreminjal. Začetki selekcije konec 18. in začetek 19. stoletja so pri mlečnem govedu izvajali selekcijo skoraj izključno na prirejo mleka, saj je bila to gospodarsko najbolj donosna lastnost. S tem so

## 1.3 zgodovina selekcije (SELEKCIJSKIH SHEM)

Začetki selekcije kot jo poznamo danes segajo v 18. stoletje. Začetnik sistematične selekcije je bil Sir Robert Bakewell (1725 – 1795), ki je pričel z načrtnim zbiranjem podatkov ter načrtnim in nadzorovanim parjenjem. Uvedel je tudi test potomcev, pri katerem testiramo skupino potomcev in na podlagi rezultatov izberemo najboljše očete. Natančni rodovniški in fenotipski podatki so tako postali temelj za razvoj učinkovite selekcije. V nadaljevanju so raziskovalci selekcijo še bolj formalizirali z uporabo statističnih metod.

### 1.3.1 Klasična napoved plemenskih vrednosti

Največji napredek se je zgodil v sredi 20. stoletja, ko je Henderson predstavil idejo plemenske vrednosti kot najboljše linearne nepristranske napovedi (*angl.* Best Unbiased Linear Prediction, BLUP) (Henderson, 1950). »Najboljše« pomeni, da je minimizirana variancanapake napovedi (angl. *prediction error variance*), »nepristranske« pa, da je povprečje ocenjenih plemenskih vrednosti za žival resnična plemenska vrednost živali. Henderson je plemenske vrednosti je napovedal po metodi mešanih modelov s sistematičnimi in živalmi kot naključnimi vplivi:

*y* =**u + Xb**+**Za**+**e; [2]** [2]

kjer je **X** matrika dogodkov, **b** je vektor ocen za sistematske vplive, **Z** je matrika dogodkov, ki povezuje naključni vpliv živali z opazovanji, **a** je vektor plemenskih vrednosti s porazdelitvijo N ~ (0, **A**\**σ2a*), e pa je vektor ostankov s porazdelitvijo N ~ (0, *σ2e*). Mešani model za napoved plemenskih vrednosti ima nekaj predpostavk. Prva je infinitezimalni model, ki pravi da **y**, **a** in **e** sledijo večrazsežni normalni porazdelitvi ter da lastnosti nadzoruje veliko število genov, ki imajo zelo majhen aditiven učinek ter so nevezani (Fisher 1918; Bulmer 1980). Druga predpostavka modela je, da poznamo genetske in okoljske variance in kovariance v populaciji osnovalcev. Tretja predpostavka je, da model lahko upošteva selekcijo in da model vsebuje vse lastnosti, na katere smo odbirali.

Napoved plemenskih vrednosti na podlagi rodovnik je povezala podatke vseh posameznikov z rodovniško matriko sorodstva **A** (Wright, 1921). Rodovniška matrika sorodstva **A** odraža pričakovano aditivno genetsko sorodstvo med posamezniki. Koeficienti v matriki **A** so dvakratniki koeficienta sorodstva (angl. *coancestry*), ki je verjetnost da dva posameznika podedujeta identičen gen po izvoru. Sama napoved plemenskih vrednosti zahteva inverzo matrike **A**. Direktno invertiranje matrike ima kubično kompleksnost, zato je Henderson (1976) razvil rekurzivno metodo za pridobitev inverze **A-1**.

Plemenska vrednost je sestavljena iz povprečja staršev (angl. *parent average*, PA) in učinka Mendelskega vzorčenja (angl. *Mendelian sampling term,* MST) (Kennedy in sod., 1988). Plemenske vrednosti lahko napovemo po modelu očeta, pri čemer uporabimo fenotipske vrednosti potomcev ter lastne fenotipske vrednosti, ali po modelu živali, kjer uporabimo fenotipske vrednosti vseh sorodnikov kot tudi lastne fenotipske vrednosti. Točnost takšne rodovniške napovedi je močno odvisna od lastne fenotipske vrednosti oz. vrednosti potomcev, narašča pa z večjo heritabiliteto in večjim številom fenotipskih vrednosti živali in njenih tesnih sorodnikov (Mrode, 2005). Poseben izziv predstavljajo lastnosti, ki so pogojene s spolom živali, npr. lastnosti mlečnosti, saj so za točno odbiro očetov potrebne fenotipske vrednosti njegovih potomk (test potomcev, tudi progeni test). Takšna selekcija je posledično zaznamovana z visoko točnostjo odbire očetov, hkrati pa tudi z dolgim generacijskim intervalom, ki omejuje genetski napredek. Plemenska vrednosti za mlade živali brez lastne fenotipske vrednosti je tako povprečje plemenskih vrednosti staršev, točnost katere je zaradi rekombinacije in segregacije majhna.

### **1.3.2 Genomska napoved plemenskih vrednosti**

Znižanje cene genotipizacije in razmah genotipizacijske tehnologije je omogočilo obsežnejšo genotipiziranje živali. Prva uporaba podatkov o genotipu za napoved plemenskih vrednosti je bila predstavljena s t.i. selekcijo s pomočjo molekularnih označevalcev (angl. *marker assisted selection*, MAS). Model za MAS poleg poligenske komponente (a) vključuje tudi sistematičen vpliv genotipa na znanih kvantitativnih vzročnih lokusih (angl. *Quantitative Trait loci,* QTL) oz. vezanih označevalcih za določeno lastnost (Fernando in Grossman, 1989). MAS ni prinesla celostne rešitve problema, saj vzročnih lokusov za določeno lastnost ne poznamo, poleg tega pa ti lokusi pojasnijo majhen delež genetske variance za lastnost. Razlog za slednje je veliko število vzročnih lokusov z majhnim vplivom, težave pri natančni oceni velikost vpliva vzročnih lokusov in dejstva, da uporabljeni genetski označevalci označevalci ponavadi niso vzročni, ampak so v vezavnem neravnovesju z vzročnim lokusom (Collard in Mackill, 2008).

Meuwissen, Hayes in Goddard (2001) so predlagali uporabo velikega števila genetskih označevalcev iz celotnega genoma za napoved t.i. genomskih plemenskih vrednosti. V modelu za napoved plemenskih vrednosti [2] rodovniško matriko sorodstva **A** nadomesti genomska matrika sorodstva **G**. Matriko **G** ustvarimo na podlagi genotipskih podatkovinodraža realizirano sorodstvo. Genomski koeficienti sorodstva bolj točno zajamejo varianco resničnega sorodstva kot rodovniški koeficienti sorodstva, saj zajamejo tudi učinek Mendelskega vzorčenja. Genomske plemenske vrednosti lahko napovemo po dveh ekvivalentnih modelih, in sicer SNP-BLUP in GBLUP. Model za SNP-BLUP je

*y* =**u + Xb**+**Zg**+**e; [**3]

kjer je **g** vektor aditivnih učinkov zamenjave alelov za uporabljene genetske označevalce, matrika **Z** pa matrika dogodkov, ki povezuje učinke označevalcev s fenotipskimi vrednostmi. Vektor **Zg** je tako vektor genomskih plemenskih vrednosti. Model za GBLUP je

*y* =**u + Xb**+**Wa**+**e; [4]**

kjer je a vektor genomskih plemenskih vrednosti, **W** pa matrika dogodkov, ki povezuje fenotipske in plemenske vrednosti. Genomske plemenske vrednosti **g** prihajajo iz porazdelitve N ~ (0, G\* *σ2a*).

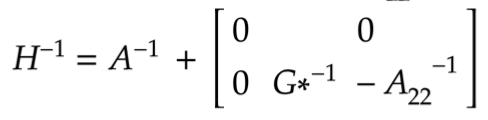
Genomske plemenske vrednosti lahko napovemo tudi z metodami Bayesovske statistike. V teh modelih modeliramo podatke na ravni podatkov samih in na ravni varianc genetskih označevalcev oz. kromosomskih segmentov (Meuwissen in sod., 2001). Vsak izmed Bayes modelov določi drugačno apriorij porazdelitev oz. mešanico porazdelitev za varianco in vplive označevalcev. Na primer, BayesA predpostavi, da imajo genetski označevalci različne variance, ki sledijo invertirani hi-kvadrat porazdelitvi, njihovi vplivi na sledijo Studentovi razdelitvi (Meuwissen in sod., 2001). BayesB sledi BayesA modelu z dodatno predpostavko, da ima le del označevalcev učinek na lastnosti, preostali označevalci pa na lastnosti na vplivajo (učinek je nič) (Meuwissen in sod., 2001). BayesC predpostavi, da imajo vsi označevalci enako varianco, učinki pa sledijo normalni porazdelitvi. Prav tako BayesC dovoljuje možnost, da π del označevalcev nima učinka (BayesCπ) (Habier in sod., 2011).

Napoved genomskih plemenskih vrednosti zahteva referenčno ali trening populacijo, ki jo sestavljajo genotipizirane živali s fenotipskimi vrednostmi. Referenčna populacija služi za oceno učinkov genetskih označevalcev, s katerimi lahko napovemo genomske plemenske vrednosti tudi mladim živalim brez fenotipske vrednosti (Meuwissen in sod., 2001). Točnost te napovedi je višja od točnosti povprečja staršev v klasični selekciji, kar je ena izmed glavnih prednosti genomske selekcije (Schaeffer, 2006; Pryce in Daetwyler, 2012). Ta točnost nam omogoča, da kot starše naslednje generacije mlade živali brez lastne fenotipske vrednosti, s čimer močno skrajšamo generacijski interval. V ZDA so v populaciji črno-bele govedi od uvedbe genomske selekcije zabeležili med 25% in 50% skrajšanje generacijskega intervala v poti selekcije očetov.

Genomska selekcija prinaša tudi spremembe v selekcijskih ciljih. Izboljša namreč učinkovitost selekcije na lastnosti, ki so težje merljive ali pa je merjenje drago, lastnosti z nizko heritabiliteto in lastnosti, ki se izrazijo pozno v življenju živali. Merjenje fenotipskih vrednosti je potrebno v referenčni populaciji, katero pa je potrebno redno obnavljati. Točnost genomskih plemenskih vrednosti narašča z večjo heritabiliteto, velikostjo referenčne populacije, številom fenotipskih vrednosti genotipiziranih živali, manjšo genetsko razdaljo med referenčno in napovedno populacijo ter z manjšanjem efektivnega števila neodvisnih kromosomskih fragmentov (Daetwyler in sod., 2008; Goddard, 2009; Habier in sod., 2010; Clark in sod, 2011; Goddard in sod., 2011). Število neodvisnih kromosomskih segmentov narekuje vzorec vezavnega neravnovesja med genetskimi označevalci in vzročnimi lokusi, kar vpliva na točnost genomske napovedi. Rekombinacija, mutacija, migracija, naključni genetski tok (angl. g*enetic drift*) in selekcija spreminjajo vzorec vezavnega neravnovesja tekom generacij, kar zmanjšuje točnost genomske napovedi. Zaradi slednjega je ključno, da referenčno populacijo redno obnavljamo z dodajanjem živali iz mlajših generacij, s čimer povečamo sorodnost in zmanjšamo genetsko razdaljo med referenčno in napovedno populacijo (Pszczola in sod., 2012).

### 1.3.3 **Združena napoved plemenskih vrednosti v enem koraku**

Leta 2009 so Legarra in sod. predstavili matriko sorodstva **H**, ki združuje rodovniško in genomsko informacijo. V matriki **H** je genomska informacija prenesena na kovarianco med negenotipiziranimi posamezniki. Čeprav je matrika **H** zapletena, ima preprosto inverzo **H-1**:

 [5]

kjer je **A22** rodovniška matrika za genotipizirane živali. Združevanje matrike **A** in **G** zahteva določene prilagoditve, saj matriki **A** in **G** nista kompatibilni. Rodovniška matrika ponavadi vključuje starejše živali, medtem ko so genotipizirane živali predvsem živali zadnjih generacij. Posledično matriki nimata iste genetske baze: rodovniška matrika predpostavlja povprečje plemenskih vrednosti nič le v populaciji osnovalcev, genomska matrika pa v celotni genotipizirani populaciji. Zaradi te razlike v obravnavani populaciji matriki **A** in **G** predpostavita tudi različno genetsko varianco za lastnost. Genomska matrika je odvisna od frekvence alelov, ki pa se zaradi naključnega genetskega toka in selekcije fiksirajo, kar zmanjša genetsko varianco. Nekompatibilnost matrik v prvem koraku rešimo z dodatkom koeficienta *a* diagonali matrike **G**, pri čemer je *a* razlika med povprečnim rodovniškim in genomskim koeficientom sorodstva med genotipiziranimi živalmi. S tem izenačimo skalo za rodovniško in genomsko napoved. V drugem koraku posodobljeno matiko **G**\* iz prvega koraka pomnožimo s koeficientom *b*, pri čemer *b*upošteva razmerje med genetsko variance trenutno in populacije osnovalcev in ponavadi manjši od 1 (Legarra in sod., 2018 – NOTES). Ko v modelu za napoved plemenskih vrednosti matriko **G** zamenjamo z matriko **H**, dobimo t.i. GBLUP v enem koraku (angl. s*ingle-step GBLUP*) (Aguilar in sod., 2010). Slednji združi vse fenotipske, rodovniške in genotipske podatke in tako dodatno poveča točnost genomske napovedi.

Kot v napovedi na podlagi rodovnika, je najbolj zahtevna operacija napovedi genomskih plemenskih vrednosti (v enem koraku) kreiranje in invertiranje matrike **G**, ki imata kubično kompleksnost s številom genotipov. Raziskovalci so zato razvili pristope, ki ne zahtevajo inverze **G-1**. Misztal in sod. (2014) so razvili t.i. algoritem za testirane in mlade (angl. *Algorithm for proven and young*, APY). Algoritem ustvari redko inverzo matrike **G-1** preko direktne inverze le malega števila živali, t.i. jedrnih (angl. c*ore*) živali,in uporabe genomske rekurzije na preostalih ne‑jedrnih (angl. n*on-core*) živalih. Drugi razširjen način pa je redukcija dimenzionalnosti matrike **G** s singularno razgradnjo vrednosti (angl. s*ingular value decomposition*) (Ødegård in sod., 2018).

Genomska selekcija je v govedoreji podvojila genetski napredek klasične selekcije navkljub nižji točnosti odbire očetov, predvsem na račun skrajšanega generacijskega intervala (Schaeffer, 2006; Wiggans in sod., 2017). Genomska selekcija je tudi povečala intenzivnost selekcije očetov, saj zaradi manjšega stroška testiranja omogoča večje število testiranih kandidatov. V Združenih državah Amerike je uvedba genomske selekcije v populaciji črno-bele govedi skrajšala generacijski interval oče-sin in oče-mati med 25 % in 50 % ter povečala selekcijski diferencial za vse lastnosti, še posebno pa za lastnosti z nizko heritabiliteto (Garcia-Ruiz in sod., 2016).

## 2 Genetska variabilnost

Selekcija dosega genetski napredek preko spreminjanja frekvenc alelov v populaciji, kar pa hkrati spreminja tudi genetsko varianco. V populacijah najdemo dva tipa procesov, ki spreminjata frekvence alelov: sistematične, med katere sodijo selekcija, migracija in mutacija, ter disperzne procese, med katere sodi naključni genetski tok in se pojavljajo v majhnih populacijah kot posledica vzorčenja (Falconer in Mackay, 1996). V rejskih programih je doseganje genetskega napredka in ohranjanje genetske variabilnosti pogosto v konfliktu. Velika intenzivnost selekcije oz. uporaba majhnega števila elitnih posameznikov zmanjšujeta genetsko variabilnost v populaciji, čeprav kratkoročno vodita do večjega genetskega napredka (Woolliams in sod., 1993). Čeprav je doseganje genetskega napredka glavni cilj selekcije, si mora ta za dolgoročni napredek prizadevati tudi za ohranjanje genetske variabilnosti. Iz enačbe za genetski napredek [1] je razvidno, da manjšanje genetske variabilnost manjša tudi genetski napredek.

Selekcija zaradi nenaključnega parjenja in večje možnosti za odbiro sorodnikov povečuje tudi parjenje v sorodstvu oz. inbriding (Woolliams in sod., 2015). Koeficient inbridinga *F* je lastnost diploidnih organizmov in je definiran kot verjetnost, da sta v posamezniku oba alela na nevtralnem mestu na genomu (brez pritiska selekcije) identična po izvoru (Wright, 1921). Inbriding povečuje homozigotnost v populaciji, kar poveča frekvenco nezaželenih genotipov (Charlesworth and Charlesworth, 1999).To vodi do depresije zaradi inbridinga, ki pomeni zmanjšanje fenotipske vrednosti za lastnosti povezane z reproduktivno sposobnostjo ali fiziološko učinkovitostjo (Falconer in Mackay, 1996).

Genetska varianca skozi generacije inbridinga pada in po *t* generacijah znaša

*σ2A,t = (1 – Ft) x σ2A,0 (5)*

(Falconer in Mackay, 1996). Vidimo, da je padec genetske variance po generacijah sorazmeren z naraščanjem inbridinga.

Inbriding lahko ocenimo na podlagi rodovnika ali genomskih informacij. Kvaliteta in točnost rodovniške ocene inbridinga močno zavisi na kvaliteti in globini rodovnika. Posledično lahko z rodovniško oceno podcenimo resničen inbriding v populaciji zaradi nepravilnega ali nepopolnega rodovnika ter predpostavke o nesorodni prvi generaciji rodovnika (npr. Lutaaya in sod., 1999; Oliehoek in sod., 2009). Inbriding lahko ocenimo tudi na podlagi genomskih podatkov, in sicer na podlagi genomskih koeficientov sorodstva ali iz deleža genoma, ki ga pokrivajo nizi homozigotnosti (angl. r*uns of homozygosity*, ROH). Genomski koeficienti inbridinga so bolj točni, saj zajamejo učinek Mendelskega vzorčenja in niso odvisni od točnosti in globine rodovnika. Prav tako lahko z nizi homozigotnosti razločimo med nedavnim in starodavnim inbridingom, pri čemer je nedaven inbriding bolj škodljiv (McQuill in sod., 2008; Ferenčaković in sod, 2012).

Čeprav je inbriding neizbežen v končno velikih populacijah, je pomembno, da uravnavamo stopnjo, s katero koeficient inbridinga narašča (*ΔF*). Upravljanje stopnje inbridinga je pomembno v dveh situacijah: prva je ohranjanje populacij, kjer je cilj minimiziranje stopnje inbridinga, druga pa selekcija, kjer je cilj čimbolj učinkovito uporabiti razpoložljivo genetsko variabilnost za ustvarjanje genetskega napredka in tako maksimirati genetski napredek ob omejitvi stopnje inbridinga (Woolliams in sod., 2015). Stopnja inbridinga določa efektivno velikost populacije (Woolliams in sod., 2015). Samo spremljanje koeficienta inbridinga *F* v starših in njihovih potomcih ni zadosten kriterij za trajnostno upravljanje s populacijo, saj se vpliv paritvenih odločitev na koeficient inbridinga izrazi z zamikom dveh generaciji, ko se ustvarijo zanke v rodovniku (Woolliams in sod., 2015). Poleg tega je *F* v populaciji relativen na bazno generacijo, za katero predpostavimo, da je neinbridirana in nesorodna. Posledično izbira bazne generacije vpliva na ocenjen *F* v populaciji (Woolliams in sod, 1998).

Tudi v populacijah govedi so raziskovalci odkrili naraščanje stopnje inbridinga kot tudi depresijo zaradi inbridinga (npr. Wiggans in sod., 1995; Thompson in sod., 1999; Doekes in sod., 2019). V populacijah govedi je povečevanje inbridinga neizbežno zaradi večgeneracijske direktne intenzivne selekcije, velike variance reproduktivnega uspeha in napovedi plemenskih vrednosti po metodi BLUP v kombinaciji s selektivno odbiro (angl. t*runcation selection*) (povzeto v Howard in sod., 2017). Uvedba genomske selekcije je močno skrajšala generacijske intervale, kar je povečalo letno stopnjo inbridinga (Doekes, 2018; Makanjuola in sod., 2020).

Za učinkovito upravljanje stopnje inbridinga moramo tako nadzorovati prispevke posameznikov preko generacij in varianco prispevkov. Idejo genetskih prispevkov sta leta 1958 razvila James in McBride. Leta 2015 jih je Woolliams tudi matematično formaliziral kot delež genov, ki jih posameznik prispeva v prihodnji genski sklad. Dolgoročni prispevki posameznikov so povezani tako z genetskim napredkom (Woolliams in sod., 1999):

ΔG = Σ*ri ai*; [6]

kjer je *ri* dolgoročen prispevek posameznika *i*, *ai* pa njegov učinek Mendelskega vzorčenja. Dolgoročni prispevki so na drugi strani povezani tudi s stopnjo inbridinga (Wray in Thompson, 1990):

ΔF = ¼ Σ*ri2.* [7]

Preko genetskih prispevkov lahko tako upravljamo tako z genetskim napredkom kot tudi inbridingom in genetsko variabilnostjo. Cilj selekcije je maksimirati genetski napredek ob vzdrževanju predhodno določene vrednosti za stopnjo inbridinga. Ker izboljšanje enega izmed teh ciljev pomeni poslabšanje drugega, je ohranjanje ravnotežja ključnega pomena.

Ta ideja je implementirana v selekciji z optimiziranimi prispevki (angl. o*ptimal contribution selection,* OCS) (Meuwissen 1997; Grundy et al. 1998; Woolliams in sod., 2015). Selekcija z optimiziranimi prispevki je nasprotje selektivni odbiri, ki le maksimizira genetski napredek. Ker je *F* posameznikov enak povprečnemu koeficientu sorodstva njihovih staršev, lahko stopnjo inbridinga nadzorujemo preko skupinskega sorodstva izbranih staršev (Woolliams in sod., 2015). Skupinsko sorodstvo staršev definiramo kot njihovo povprečno sorodstvo uteženo z njihovimi prispevki k naslednji generaciji. Za samo optimizacijo so potrebne dodatne omejitve, ki izhajajo iz praktičnega vidika parjenja: zagotoviti moramo, da so prispevki vseh posameznikov ne-negativni, da je vsota moških kot tudi ženskih posameznikov ½ ter da je vsota prispevkov vseh posameznikov enaka 1 (Woolliams in sod., 2015).

Na drugi strani lahko z migracijo povečamo genetsko variabilnost določene populacije. V populacijah govedi to vršimo preko uvoza genetskega materiala predvsem očetov. Vendar je v populacijah prisotna interakcija genotip-okolje, ki lahko vodi do prerangiranja posameznikov med okolji. Prerangiranje pogosto merimo z genetsko korelacijo za lastnosti (Falconer in Mackay, 1996), pri čemer korelacija manj kot ena nakazuje prisotnost interakcije genotip-okolje. Pri tem je dolgoročno sodelovanje in izmenjava živali med rejskimi programi možna le, če je korelacija med lastnostjo v obeh populacijah med 0.8 in 0.9 (Mulder in Bijma, 2006). Učinek sodelovanja na večanje genetskega napredka in manjšanje stopnje inbridinga je večji v genomski selekciji, prav tako genomska selekcija dovoljuje manjše korelacije med okolji za dolgoročno sodelovanje populacij (Cao in sod., 2020).

## 1.3 STOHASTIČNE SIMULACIJE

Selekcija je dolgotrajen in drag postopek z mnogo spremenljivkami, vpliva katerih še ne poznamo. Slednje onemogočajo deterministične izračune genetskih parametrov, kot so genetski napredek in genetska variabilnost. Testiranje različnih modelov v praksi je zaradi dolgih generacijskih intervalov, števila živali in potrebnih infrastrukturnih kapacitet nemogoče. Računalniške simulacije nam omogočajo hitro, poceni in ponovljeno testiranje rejskih programov, statističnih metod, predpostavk kvantitativnih modelov, analizo kratko-, srednje- in dolgoročnih posledic selekcije. Z računalniškimi simulacijami lahko poustvarimo tako naključne kot sistematične procese – od kreiranja nove generacije posameznikov, do selekcije in parjenja izbranih staršev.

### 1.3.1 Simulacija genoma

Simulacija genomov in vzročnih mehanizmov je izjemno kompleksna zaradi različnih virov variabilnosti, različnih možnih zgodovin populacije ter negotovosti glede stopnje rekombinacije, stopnje mutacije in načina ter distribucije učinkov genov (Daetwyler in sod., 2013). Pri simulaciji genoma moramo upoštevati genetske parametre simuliranega organizma oz. vrste, saj se zaradi različne demografije njihovi genetski parametri razlikujejo. Pri simulaciji genoma moramo tako upoštevati velikost in število kromosomov, efektivno velikost populacije prednikov, vzorec vezavnega neravnovesja, stopnje mutacij ter stopnje rekombinacij. Na podlagi slednjih parametrov simuliramo genotipe oz. haplotipe bazne populacije. V praksi so se uveljavile tri glavne metode simulacije genomov. Prva je metoda na podlagi ponovnega vzorčenja (angl. m*ethods based on resampling*), ki uporabi resnične podatke genomskega zaporedja ter rodovnik za simulacijo genomov trenutne populacije. Ta metoda najbolje posnema resnične frekvence alelov in vezavno neravnovesje, ne omogoča pa vpeljave novih mutacij ali genetskih značilnosti. Druga je metoda na podlagi koalescentne teorije, s katero simuliramo genom nazaj v čas. Metoda identificira in združi (angl. *Coalesce*) skupne prednike danega vzorca neznanih genotipov s stohastičnim procesom, pri čemer upošteva mutacijo, rekombinacijo in migracijo. Tretja metoda je metoda »naprej v čas«, ki je tudi najlažja za implementacijo in najbolj fleksibilna. Ta metoda ustvarja populacijo s časom glede na navedene genetske in demografske parametre.

### 1.3.2 Simulacija fenotipa

Za simulacijo genetskih in fenotipskih vrednosti moramo simulirati delovanje genov. Izmed simuliranih lokusov izberemo množico z učinkom, učinke pa vzorčimo iz izbrane porazdelitve. Porazdelitev učinkov je lahko enostavna, npr. normalna, lahko pa je zelo kompleksna ter vključuje nelinearne učinke. Naprednejši simulatorji omogočajo določitev enostavnih lastnosti z aditivnimi učinki kot tudi lastnosti z učinki dominance, epistaze in interakcije genotip-okolje (Gaynor in sod., 2020). Na podlagi učinkov genov izračunamo pravo genetsko vrednost posameznikov, fenotipske vrednosti pa simuliramo z dodatkom naključnega odklona za ostanek h genetski vrednosti. Odkloni ostankov so vzorčeni iz normalne porazdelitve s povprečjem vplivov 0 in varianco, ki je enaka varianci za ostanek. Varianca za ostanek je določena na način, da se doseže nastavljena heritabiliteta za lastnost.

### 1.3.3 Simulacija populacije

Iz osnovnega nabora haplotipov s ponavljanjem vzorčimo haplotipe za prvo generacijo populacije, pri čemer upoštevamo naključna procesa rekombinacije in segregacije. Simulaciji genoma in posameznikov v osnovni populaciji sledi odbira staršev in parjenje. Začetno velikost in strukturo populacije vzpostavimo s t.i. izgorevanjem (angl. *burn-in*), kjer določeno število generacij izvajamo naključna izbira staršev in parjenje. Temu sledi simulacija selekcije z načrtno odbiro staršev. Kriterij za odbiro staršev je lahko povprečje staršev, klasična plemenska vrednost, genomska plemenska vrednosti, prava genetska vrednost ali fenotipska vrednost. Odbiri staršev sledi parjenje in ustvarjanje novorojenih posameznikov ter njihovih genomov, pri čemer pride do rekombinacije in segregacije starševskih gamet (Faux in sod., 2016).

Simulacijski programi omogočajo simulacijo rejskih programov različnih kompleksnosti. Simuliramo lahko enostavne rejske programe z diskretnimi generacijami in malim številom rejskih odločitev (npr. samo odbira staršev in zadnje generacije). Lahko pa simuliramo tudi kompleksne rejske programe s prekrivajočimi generacijami, velikim številom rejskih odločitev, večstopenjskim testiranjem živali itd. Na ta način lahko posnemamo resnične rejske programe. V simulaciji lahko uporabimo tudi selekcijske parametre povzete po dotičnem rejskem programu, kot npr. število novorojenih živali, število živali v testu potomcev, število odbranih bikov itd. To nam omogoča simulacijo specifične resnične populacije s specifičnimi genetskimi parametri in testiranje hipotez na sistemu, ki je primerljiv z našim tarčnim resničnim sistemom.

4 MAJHNE POPULACIJE

Majhne populacije so zaznamovane z različnimi genetskimi, selekcijskimi in praktičnimi posebnostmi ter omejitvami. Te so predstavljale izziv že v klasični selekciji, v genomski selekciji pa so bolj izrazite ter hitro povečujejo razkol med majhnimi in večjimi populacijami. Čeprav je genomska selekcija že nekaj let uveljavljena tehnologija v večjih populacijah, se majhne populacije še vedno spopadajo z izzivi tako pri inicializaciji kot tudi kontinuiranem izvajanju genomske selekcije. Omejitve majhnih populacij se izražajo v doseganju genetskega napredka in ohranjanju genetske variabilnosti, izhajajo pa iz pomanjkanja tako živali, trajnostne strategije kot tudi finančnih sredstev

Majhne populacije so omejene pri doseganju genetskega napredka. Omejitve so v veliki meri vezane na manjše število živali, ki omejuje i) intenzivnost selekcije in ii) točnost klasične in genomske selekcije.

i) Majhno število živali pomeni manjše število testiranih selekcijskih kandidatov in posledično manjšo intenzivnost selekcije. Slednje se je še posebno izrazilo z uvedbo genomske selekcije, saj ta omogoča cenejše in hitrejše testiranje velikega števila selekcijskih kandidatov.

ii) V klasični selekciji majhne populacije dosegajo manjše točnosti testa potomcev zaradi manjšega števila potomcev s fenotipskimi vrednosti. Zaradi slednjega in majhnega števila progeno testiranih bikov imajo majhne populacije težave tudi s sestavo zadostne referenčne populacije za genomsko napoved. To zaradi manjše točnosti napovedi omejuje prednost genomske pred klasično selekcijo (Thomasen in sod., 2014).

Težavo nezadostne referenčne populacije lahko rešujemo s sodelovanjem v mednarodnih konzorcij in uporabo mednarodne enopasemske referenčne populacije (Jorjani, 2012; Špehar in sod., 2013; Schöpke in Swalve, 2016) ali pa uporabo večpasemske referenčne populacije. Uspeh slednjih zavisi od genetske sorodnosti populacij ali pasem in o podobnosti vzorcev genetskega neravnovesja, Oba pristopa povečanja referenčne populacije v majhnih populacijah povečata točnost genomske selekcije, vendar je združevanje znotraj pasme bolj učinkovito kot združevanj pasem (Hozé in sod., 2014; Lund in sod., 2014). Alternativa referenčni populaciji progeno testiranih bikov je tudi referenčna populacija ženskih živali. Zaradi manjše količine fenotipskih podatkov na kravo v primerjavi s progeno testiranimi biki je relativen prispevek ene krave k referenčni populaciji manjši kot enega progeno testiranega bika (de Roos, 2011). Vendar so raziskave pokazale, da lahko z dodatkom ženskih podatkov v referenčno populacijo dosežemo konkurenčnost majhnih populacij (Thomasen in sod., 2014). V primerjavi s samo moško referenčno populacijo dodatek ženski poveča točnost napovedi ter zmanjša stopnjo inbridinga. Raziskave so pokazale tudi, da referenčna populacija hčera s fenotipskimi podatki omogoča večjo točnost napovedi kot referenčna populacija njihovih očetov (Buch in sod., 2012) ter da lahko točnost genomske napovedi povečamo s genotipizacijo živali z lastnimi fenotipskimi vrednostmi (van Grevenhof in sod., 2012).

Rejski programi v majhnih populacijah morajo posvečati posebno pozornost tudi ohranjanju genetske variabilnosti. Ta zagotavlja dolgoročen genetski napredek ter s tem ohranja konkurenčnost populacij in upravičuje obstoj nacionalnih rejskih programov. V majhnih populacijah je učinek naključnega genetskega toka močnejši in ima za posledico naključno nihanje frekvenc alelov. To nihanje lahko privede do naključne fiksacije ali izgube alela ter povečane homozigotnosti in inbridinga (Falconer in Mackay. 1996). Majhne populacije lahko tudi vnašajo genetsko variabilnost preko uvoza genetskega materiala., s čimer zmanjšajo tudi stopnjo inbridinga. Uvoz povečuje sorodnost znotraj pasme oz. globalne populacije govedi, vendar pa znižuje sorodnost znotraj majhnih populacij. Majhne populacije uvažajo očete tudi za večanje genetskega trenda, saj so ti genetsko superiorni zaradi bolj učinkovite selekcije v večjih populacijah. Vendar z uvozom tudi manjšamo pomen domače selekcije.

Majhne populacije morajo tako oblikovati trajnostno strategijo, ki bo omogočila i) učinkovito doseganje genetskega napredka za konkurenčnost na mednarodnem trgu; ii) učinkovito upravljanje z genetsko variabilnostjo, ki bo omogočala tako kratko- kot tudi dolgoročni genetski napredek; iii) ekonomsko učinkovitost.

Ker si omenjeni cilji nasprotujejo, je njihovo uravnoteženje težavno za vsak rejski program, še posebno pa v majhnih populacijah. Strategije, ki bi maksimirale genetski napredek pogosto niso optimalne z vidika ohranjanja genetske variabilnosti ali pa ekonomske učinkovitost. Na primer, da bi ohranile konkurenčnost na mednarodnem trgu, morajo majhne populacije najprej nadoknaditi zaostanek v genetskem napredku. Kratkoročno je to najlažje doseči s povečanjem intenzivnost selekcije, ki pa je z vidika genetske variabilnosti problematično zaradi že tako majhnega števila razpoložljivih staršev. Prav tako bi genetski napredek maksimirali z uvedbo genomske selekcije z veliko referenčno populacijo, vendar to z ekonomskega vidika ni najbolj učinkovita strategija. Zaradi pomanjkanja infrastrukture in opreme morda z ekonomskega vidika tudi ni optimalno spremljanje vseh lastnosti in rodovnika (Biscarinni in sod., 2015).

Za dolgoročno konkurenčnost pa morajo majhen populacije čimprej uvesti genomsko selekcijo. Prvi problem je tako, kako inicializirati genomsko selekcijo in z razpoložljivimi sredstvi sestaviti referenčno populacijo za genomsko napoved. Poleg trajnostne strategije v smislu doseganja selekcijskih ciljev mora biti dolgoročno program tudi ekonomičen.

# RAZPRAVA IN SKLEPI

# RAZPRAVA

Majhne populacije so pri izvajanju selekcije omejene zaradi genetskih in praktičnih posebnosti. Zaradi majhnega števila živali ter omejenih finančnih sredstev za testiranje selekcijskih kandidatov ne morejo izvajati zelo intenzivne selekcije. Prav tako majhno število živali in potomcev po očetu omejuje točnost tako klasične kot tudi genomske selekcije. Omejitve glede števila živali in točnosti vplivajo tudi na genetsko variabilnost v populaciji. Navkljub vsem posebnostim morajo majhne populacije zagotavljati tako kratko- kot dolgoročni genetski napredek, da ostanejo konkurenčne na mednarodnem trgu in upravičijo obstoj domače selekcije.

Genomska selekcija je uveljavljena tehnologija v velikih populacijah. Nasprotno se veliko majhnih populacij še vedno spopada z njeno uvedbo, predvsem zaradi pomanjkanja finančnih sredstev, skepticizma uporabnikov in pomanjkanja strokovne podpore. Neuvedba genomske selekcije povečuje razkol med velikimi in majhnimi populacijami. Zaradi razlike v genetskem napredku majhne populacije tako povečujejo uvoz tujega genetskega materiala in s tem zmanjšujejo pomen domače selekcije.

Klasična in genomska selekcija se razlikujeta tudi v načinu doseganja genetskega napredka in financiranja (prioritet) rejskih akcij. Klasični program implementira progeni test. Ta zaradi visokih stroškov omejuje intenzivnost selekcije in ima za posledico dolg generacijski interval, hkrati pa omogoča visoko točnost odbire. Slednja je tako glavno gonilo napredka v klasični selekciji. Rejski programi s klasično selekcijo večino finančnih sredstev tako namenijo zbiranju fenotipskih vrednosti, ki je nerazdružljivo povezano s progenim testom. Genomska selekcija pa dosega genetski napredek predvsem preko močno skrajšanega generacijskega intervala, večje intenzivnosti selekcije moških živali in večje točnosti odbire mladi živali brez lastnih fenotipskih vrednosti. Zbiranje fenotipskih podatkov je še vedno ključnega pomena za obnavljanje referenčne populacije, vendar pa ni več intrinzično povezano s selekcijo. Genomska selekcija mora tako uravnotežiti investiranje v zbiranje fenotipskih podatkov in genotipizacijo živali za obnavljanje referenčne populacije in selekcijo. Razlika v investicijskih prioritetah med obema programoma predstavlja problem za prehod na genomsko selekcijo, še posebno v rejskih programih z omejenimi sredstvi ali pa v državah, kjer so različne rejske aktivnosti financirane iz različnih virov.

Majhne populacije morajo tako stremeti k visoki učinkovitosti selekcije v vseh pogledih. V nadaljevanju smo tako razpravo razdelili na: i) učinkovitost v doseganja genetskega napredka; ii) učinkovitost ohranjanja genetske variabilnosti; in iii) ekonomske učinkovitosti.

Genetski napredek

Da bi ohranile konkurenčnost na mednarodnem trgu, morajo majhne populacije dosegati genetski napredek primerljiv z velikimi populacijami oz. globalnim genetskim napredkom. Če hočejo ujeti genetski napredek velikih populacij, morajo zaostriti enega od parametrov, ki povečujejo genetski napredek: intenzivnost, točnost ali pa dodatno skrajšati generacijski interval. V naši raziskavi (poglavje X.x) smo s stohastično simulacijo testirali vpliv števila odbranih očetov in leta uporabe na genetski napredek.