

*Bioinformacijski pristopi k analizi kompleksnih  
lastnosti pri domačih živalih*

Eva Čeh

DOKTORSKA DISERTACIJA

PREDANA

FAKULTETI ZA RAČUNALNIŠTVO IN INFORMATIKO

KOT DEL IZPOLNJEVANJA POGOJEV ZA PRIDOBITEV NAZIVA

DOKTOR ZNANOSTI

S PODROČJA

BIOZNANOSTI, ŠTUDIJSKE SMERI BIOINFORMATIKE



Ljubljana, 2013



## IZJAVA

*Izjavljam, da sem avtorica dela in da slednje ne vsebuje materiala, ki bi ga kdorkoli predhodno že objavil ali oddal v obravnavo za pridobitev naziva na univerzi ali na drugem visokošolskem zavodu, razen v primerih, kjer so navedeni viri.*

— Eva Čeh —  
december 2013

ODDAJO SO ODOBRILI

dr. Peter Dovč

*redni profesor za področje molekularne genetike in biotehnologije živali*

MENTOR IN ČLAN OCENJEVALNE KOMISIJE

dr. Tomaž Curk

*docent za področje računalništva in informatike*

PRESEDNIK OCENJEVALNE KOMISIJE

dr. Gregor Majdič

*redni profesor za področje veterinarske fiziologije*

ZUNANJI ČLAN OCENJEVALNE KOMISIJE



## PREDHODNA OBJAVA

Izjavljam, da so bili rezultati obravnavane raziskave predhodno objavljeni/sprejeti za objavo v recenzirani reviji ali javno predstavljeni v naslednjih primerih:

- [1] E. Čeh, P. Dovč. Population structure and genetic differentiation of livestock guard dog breeds from the Western Balkans. V: *Journal of Animal Breeding and Genetics*, v tisku.
- [2] E. Čeh, P. Dovč. Identifikacija kandidatnih lokusov za ektopični ureter pri kraškem ovčarju. V: *2. kolokvij iz genetike, Piran, 16. september 2011*, str. 136, Ljubljana, 2011. Slovensko genetsko društvo.
- [3] E. Čeh, A. Islas-Trejo, J.F. Medrano, A. Razpet, G. Rincon, P. Dovč. SNP association analysis of ureteral ectopia in Karst Shepherds. V: *9th Congress of the Slovenian Biochemical Society, Maribor, 12th - 15th October 2011. Abstract book*, str. 155, Maribor, 2011. Zavod za zdravstveno varstvo.
- [4] E. Čeh, P. Dovč. Ectopic ureter in the Karst Shepherd breed. V: *25. simpozij o aktualnih boleznih malih živali, 30. - 31. marec 2012, Portorož. Zbornik referatov*, str. 66–70, Ljubljana, 2012. Slovensko združenje veterinarjev za male živali.
- [5] E. Čeh, A. Razpet, P. Dovč. Genetic issues of a rare breed: a case of a Karst Shepherd. V: *The 6th International Conference on Advances in Canine and Feline Genomics and Inherited Diseases : May 28 - June 1, 2012 [abstracts]*, str. 79. Visby, Švedska, 2011.
- [6] E. Čeh, A. Razpet, P. Dovč. Population parameters in the Karst Shepherd. V: *Genetika 2012 : book of abstracts = zbornik povzetkov*, str. 193, Ljubljana, 2012. Slovensko genetsko društvo.

Potrjujem, da sem pridobila pisna dovoljenja vseh lastnikov avtorskih pravic, da lahko zgoraj navedene vire vključim v pričujočo disertacijo. Potrjujem, da zgoraj navedeni viri opisujejo rezultate raziskav, izvedenih v času mojega podiplomskega študija na Univerzi v Ljubljani.



*Mami in očetu, ki sta me naučila ljubezni do dela, učenja in poštenosti.*



## POVZETEK

V pričujoči disertaciji so opredeljeni genski mehanizmi, odgovorni za kompleksne lastnosti pri domačih živalih z uporabo različnih bioinformacijskih pristopov. Kompleksne so običajno tiste lastnosti, na katere vpliva večje število genov, ki med seboj delujejo vzajemno, nanje pa vplivajo tudi okoljski dejavniki. V nasprotju s preprostimi, mendelskimi oziroma enogenskimi lastnostmi klasične genetike so vzorci dedovanja kompleksnih lastnosti zaradi vpletenosti več dejavnikov težavnejši za razumevanje in raziskovanje, a so kompleksne lastnosti v resnici bolj pogoste. Zahvaljujoč hitremu napredku razvoja visokozmogljivostnih tehnologij sekvenciranja in genotipizacij lahko hitro zberemo ogromne količine genomskih podatkov, nov izziv pa je način, kako iz zbrane kopice podatkov učinkovito razčleniti biološko pomembne informacije o preučevanih fenotipskih lastnostih.

Disertacija sestoji iz treh študijskih primerov: populacijsko-genetske raziskave, analize genske bolezni ter analize transkriptoma laboratorijskega modela kužne bolezni. V vseh treh raziskavah smo poskušali, čeprav na različnih ravneh genoma, odkriti genske dejavnike ter opisati in razložiti kompleksne genske vzorce, ki vplivajo na preučevane lastnosti. V tem delu naslavljam biološke probleme in pri tem izkoriščamo različne računske pristope. V okviru populacijsko-genetske raziskave na primeru pastirskih pasem psov razkrivamo presenetljivo močan vpliv genskega zdrsa na sliko genskih odnosov znotraj majhnih populacij. V naslednjem poglavju opisujemo, kako smo omejen genski sklad pasme kraškega ovčarja izkoristili za preučevanje uhajanja seča, asociacijsko preiskavo celotnega genoma ter kartiranje homozigotnosti ter katere sugestivne genomske lokuse za nagnjenost k ektopiji sečevodov pri tej pasmi smo odkrili. V okviru tretjega raziskovalnega primera se posvečamo eksperimentalnemu modelu kužne agalaksije ter prikazujemo časovne vzorce transkriptoma patogenega in gostiteljevega organizma v prvih 24 urah okužbe na osnovi podatkov sekvenciranja RNA. V okviru

vseh treh preučevanih primerov smo odkrili kandidatne vzročne dejavnike, ki ponujajo vpogled v specifične populacijske oziroma fenotipske lastnosti, pa tudi kompleksne lastnosti na sploh ter predlagali iztočnice za prihodnje raziskave ter uporabo pridobljenega znanja v rejski praksi.

*Ključne besede:* kompleksne lastnosti, bioinformacijska analiza, domače živali, sekvenciranje RNA, preiskava celotnega genoma, populacijska genetika

## ABSTRACT

In this thesis, we focus on the characterization of the genetic mechanisms responsible for the complex traits in domestic animals using a variety of bioinformatics approaches. Complex traits are usually considered to be affected by a number of interacting genes, as well as by environmental factors. In contrast to simple Mendelian traits, inheritance patterns of complex traits are more difficult to study and understand, but are in reality more common. Thanks to the rapid progress of the development of high-throughput sequencing and genotyping technologies, huge amounts of genomic data can be quickly gathered. Therefore we are now faced with a new challenge of interpreting huge collections of genomic data and dissection of biologically relevant information.

This dissertation consists of three case studies: population genetics study, analysis of a genetic disease and a transcriptomic analysis of a laboratory model of an infectious disease. In all three studies we attempt to identify the underlying genetic factors and interpret the complex genetic patterns that affect the studied traits. We address biologically relevant problems using different computational approaches. The population genetics study of livestock guard dog breeds from the Western Balkans reveals a surprisingly strong influence of genetic drift on the resolution of genetic relationships of studied breeds. In the next section we take advantage of a limited gene pool of the Karst Shepherd breed for the purpose of a genome-wide scan and homozygosity mapping of ureteral ectopia. We further present suggestive genomic loci for susceptibility to ureteral ectopia in this breed. In the third case study we investigate the experimental model of contagious agalactia based on RNA sequencing data and show temporal transcriptomic profiles of the pathogen and host organisms during the first 24 hours of infection. We identify possible causal genetic factors in all three study cases, offer insight into the specific population characteristics or phenotypic traits as well as com-

plex traits in general, and also lay foundation for future research and implementation of the acquired knowledge in breeding practices.

*Key words:* complex traits, bioinformatics analysis, domestic animals, RNA-Sequencing, genome-wide association study, population genetics

## ZAHVALA

*Hvala prof. dr. Petru Dovču, doc. dr. Tomažu Curku, prof. dr. Gregorju Majdiču, prof. dr. Juanu F. Medranu, dr. Gonzalu Rinconu, Almi D. Islas-Trejo, dr. Andreju Razpetu, Mojci Emeršič, prof. dr. Aleksandri Domanjko-Petrič, doc. dr. Darji Pavlin, dr. Minji Zorc, dr. Milivoju Uroševiću, Stanku Pašiču, prof. dr. Inu Čuriku, prof. dr. Blažu Zupanu, članom društva ljubiteljev in vzrediteljev kraških ovčarjev (DLVKOS), kolegom in sodelavcem s Katedre za genetiko, imunologijo in animalno biotehnologijo in premnogim pasjim lastnikom, ki so mi pomagali zbrati potreben genski material.*

— Eva Čeh, Ljubljana, december 2013.



# KAZALO

<i>Povzetek</i>	<i>i</i>
<i>Abstract</i>	<i>iii</i>
<i>Zahvala</i>	<i>v</i>
<b>1</b> <i>Uvod</i>	<b>1</b>
1.1 Motivacija . . . . .	2
1.2 Prispevki k znanosti . . . . .	2
1.3 Pregled vsebine . . . . .	4
<b>2</b> <i>Populacijska genetika pastirskih pasem psov Zahodnega Balkana</i>	<b>5</b>
2.1 Ozadje . . . . .	6
2.2 Metode in rezultati . . . . .	7
2.2.1 Vzorci in izolacija DNA . . . . .	7
2.2.2 Genska pestrost na osnovi mikrosatelitnih označevalcev . . . . .	7
2.2.3 Genska pestrost na osnovi mitohondrijske DNA . . . . .	10
2.2.4 Analiza rodovnikov . . . . .	13
2.2.5 Razlikovanje pasem . . . . .	17
2.2.6 Populacijska struktura . . . . .	19
2.3 Diskusija in prihodnji izzivi . . . . .	22
<b>3</b> <i>Odkrivanje genskega ozadja ektopije sečevodov pri kraškem ovčarju</i>	<b>25</b>
3.1 Ozadje . . . . .	26
3.2 Metode in rezultati . . . . .	27
3.2.1 Izjava o etiki . . . . .	27

3.2.2	Fenotipski podatki in analiza rodovnikov . . . . .	27
3.2.3	Asociacijska preiskava celotnega genoma in kartiranje homozigotnosti . . . . .	30
3.3	Diskusija in prihodnji izzivi . . . . .	37
4	<i>Analiza transkriptoma celic mlečne žleze kože po okužbi z bakterijo Mycoplasma agalactiae in vitro</i> . . . . .	41
4.1	Ozadje . . . . .	42
4.2	Metode in rezultati . . . . .	43
4.2.1	Opis eksperimentalnih podatkov . . . . .	43
4.2.2	Kartiranje zaporedij na referenčni genom . . . . .	44
4.2.3	Določitev transkriptoma s pristopom <i>de novo</i> . . . . .	47
4.2.4	Diferencialno izražanje genov . . . . .	48
4.2.5	Analiza funkcij diferencialno izraženih genov . . . . .	52
4.3	Diskusija in prihodnji izzivi . . . . .	60
5	<i>Zaključek</i> . . . . .	63
	<i>Literatura</i> . . . . .	67

*Uvod*

## *1.1 Motivacija*

Do nedavnega je bila genetika ozko usmerjena, skoraj nekoliko ezoterična znanost. Njen domet je bil omejen na iskanje mutacij z močnimi učinki, ki so prisotne pri manjšem številu osebkov in povzročajo redke lastnosti. Še desetletje nazaj je bilo poznanih le malo genskih različic, ki so v populacijah pogoste in vplivajo na lastnosti, ki ne sledijo preprostim mendelskim vzorcem dedovanja in jih ne opredeljujejo aleli le enega genskega lokusa. Takšne lastnosti pojmujejo kot kompleksne. Večina fenotipov, ki so pomembni v medicinskem, biološkem ali ekonomskem smislu, kot so npr. življenjska doba, telesna teža ter višina ali tveganje za sladkorno bolezen, je kompleksnih. Nekateri od dejavnikov, ki vplivajo na kompleksnost lastnosti, so nepopolna penetranca, večgensko dedovanje, medsebojni vplivi genov in vplivi okolja. Da bi bolje spoznali vzroke in posledice fenotipske raznolikosti v populacijah, predvidevali nagnjenost k pogostim boleznim in gojili rastline ter vzrejali živali, s katerimi bi lahko izpolnjevali potrebe sveta, je ključno razumevanje kompleksnih lastnosti.

Zakaj so genske različice, ki vplivajo na pojav bolezni, prisotne v populaciji? Kako populacijska struktura vpliva na razlikovanje odnosov med sorodnimi populacijami? Koliko lokusov in kakšni genski mehanizmi vplivajo na dano lastnost? Kakšni so učinki teh lokusov? Katere biološke poti so odgovorne za pojav kompleksnih lastnosti? To so samo nekatera od vprašanj, ki si jih zastavljamo v tem doktorskem delu in empirično, z uporabo različnih bioinformatičnih pristopov, odkrivamo odgovore na osnovi treh študijskih primerov: raziskave populacijske strukture in genskih odnosov med pastirskimi pasmami psov Zahodnega Balkana, raziskave, usmerjene k odkrivanju genskih dejavnikov, ki vplivajo na pojav ektopije sečevodov pri kraškem ovčarju ter analize transkriptoma celičnega modela medsebojnih vplivov med patogenom in gostiteljem med kužno agalakcijo.

## *1.2 Prispevki k znanosti*

V disertaciji so predstavljeni naslednji prispevki k znanosti:

- Ocena genske pestrosti pastirskih pasem psov Zahodnega Balkana na osnovi mikrosatelitnih označevalcev in mitohondrijske DNA (mtDNA). Z gensko pestrostjo, ocenjeno na osnovi mikrosatelitnih označevalcev, smo določili prispevke k celokupni pestrosti ter prispevke k pasemski oziroma medpasemski raznolikosti. Gensko pestrost pri kraškem ovčarju smo ocenjevali tudi glede na časovni vidik.

Raznolikost D-zanke mtDNA smo analizirali v širšem smislu pastirskih pasem, določili porazdelitve haplotipov v pasmah, odkrili dva nova haplotipa in potrdili genski tok med kraševcem in šarplanincem.

- Analiza rodovnikov. Določili smo ustanovitelje trenutne populacije kraškega ovčarja ter njihove prispevke glede na pasmo ali izvor. Določili smo tudi populacijske parametre kot npr. efektivno velikost populacije, koeficient inbreedinga, efektivno velikost ustanoviteljev, ki potrjujejo hipotezo o pomembni vlogi genskega zdrsa pri tej pasmi.
- Razlikovanje pastirskih pasem psov Zahodnega Balkana in njihova populacijska struktura. Kljub skupni preteklosti sorodstveni odnosi med nekaterimi pasmami danes niso več očitni (kraški ovčar - šarplaninec), po drugi strani pa je bilo nekatere druge pasme v naši raziskavi gensko težje razločiti (šarplaninec - kavkaški ovčar).
- Določitev nekaterih lastnosti fenotipa ektopije sečevodov in analiza dedovanja po rodovniku kraškega ovčarja. V triletnem obdobju smo zbirali diagnostične podatke o tej razvojni bolezni pri kraškem ovčarju ter tako opredelili nekatere tipične značilnosti. Z rodovniško analizo smo predlagali kompleksen vzorec dedovanja.
- Asociacijska preiskava celotnega genoma in kartiranje homozigotnosti za ektopijo sečevodov pri kraškem ovčarju. Odkrili smo kandidatne gene za ekopijo sečevodov ter jih potrdili s pregledom mendelskih neskladij in določitvijo genotipov pri drugih pasmah, ki nimajo visoke pogostosti ekopičnega sečevoda. Njihovo vlogo pri razvoju sečil smo potrdili z analizo funkcij genov.
- Določitev transkriptoma celic mlečne žleze kože med okužbo z bakterijo *Mycoplasma agalactiae* s strategijami, ki uporabljajo referenčni genom in strategijami, ki ga ne. S kombiniranim pristopom smo izkoristili prednosti obeh načinov.
- Odkritje možnih novih fragmentov prepisov, ki lahko služijo za strukturno anotacijo genoma kože.
- Odkritje diferencialno izraženih genov med okužbo celičnega modela kužne agalaksije ter odkritje potencialnih virulentnih dejavnikov *M. agalactiae*.

- Določitev časovnih vzorcev izražanja ter določitev bioloških poti s spremenjenim uravnavanjem med okužbo.

### *1.3 Pregled vsebine*

V disertaciji raziskujemo kompleksne lastnosti domačih živali na treh študijskih primerih. V vsakem od poglavij študijskih primerov v posameznih odsekih predstavljamo ozadje, metode in rezultate ter diskusije s prihodnjimi izzivi. V poglavju 2 predstavljamo populacijsko genetiko kraškega ovčarja, šarplaninca in tornjaka ter raziskujemo genske mehanizme, ki vplivajo na njihove sorodstvene odnose. Poglavje 3 prikazuje postopek in rezultate odkrivanja genske nagnjenosti k ektopičnemu sečevodu pri kraškem ovčarju. V poglavju 4 preučujemo različne pristope za odkrivanje transkriptoma in opazujemo časovne vzorce izražanja genov med okužbo celičnega modela kužne agalaksije. Poglavje 5 s pregledom znanstvenih prispevkov povzema zaključke doktorskega dela in ponuja smernice za prihodnje raziskave in dejavnosti v rejski praksi.

*Populacijska genetika  
pastirskih pasem psov  
Zahodnega Balkana*

## 2.1 Ozadje

Pes (*Canis familiaris*) je eden najboljših modelov za raziskave znotrajvrstne genske pestrosti, saj velja za fenotipsko najbolj raznolikega sesalca. Veliko število pasem omogoča medsebojne primerjave, saj je možno večino rodovniških psov na osnovi genskih podatkov razvrstiti glede na pasmo [1–3]. To je posledica nedavnega izvora pasem psov, ki so nastale iz omejenega števila ustanoviteljev (angl. founders) in predstavljajo zaprte genske sklade z majhnimi učinkovitimi populacijskimi velikostmi. Glede na podatke mikrosatelitov [2, 4], SNP-jev [5] in CNV-jev [6] je možno različne pasme psov razvrščati tudi glede na telesno zgradbo (konformacijo) oz. pasemski tip. Na razlikovanje (diferenciacijo) sodobnih pasem psov poleg njihovih zgodovinskih sorodnosti vplivajo tudi sedanje rejске prakse. Delovne pasme psov se na primer uvrščajo bližje vaškim psom mešancem kot pa pasmam, pri vzreji katerih je v ospredju telesna zgradba [7]. Načrtno parjenje (angl. assortative mating) ustvarja pasemske podskupine, ki se ne ujemajo s pasemskimi podtipi, kot jih prepoznavao kinološke organizacije. Slednjih podtipov v nekaterih primerih (npr. velikostne skupine pudljev [8] ali šnavcerjev [9]) gensko ni možno razlikovati. Algoritmi za razvrščanje v skupine (angl. clustering) prav tako vedno ne razlikujejo niti psov sorodnih pasem (npr. mali in veliki vendejski ostrodlaki baset [4], bulterier in pritlikavi bulterier [5]). Nekatere raziskave razkrivajo tudi združevanje nesorodnih pasem (npr. shih-tzu v skupini lovskih pasem v [2]), kar zbuja vprašanja o utemeljenosti takšnega razvrščanja. Prav tako so veje v filogenetskih drevesih, ki razlikujejo pasme, običajno kratke, podpora s samovzorčenjem (angl. bootstrap) pa nizka, kar je skladno z mrežno evolucijo različnih pasem psov, na katero pomembno vplivajo medpasemska križanja in populacijska nihanja, ki otežujejo razkrivanje odnosov med pasmami.

Pastirske pasme psov Zahodnega Balkana so dober primer domačih živali s kompleksnim vzorcem genske pestrosti, čeprav so fenotipsko le velikostne in barvne različice pastirskega tipa psa, ki ima dolgo zgodovinsko prisotnost v tem predelu sveta [10–12]. Tornjak, bosansko-hercegovski hrvaški pastirski pes, je bil priznan s strani Mednarodne kinološke zveze (FCI) šele v letu 2007. Predniki kraških ovčarjev in šarplanincev so bili vodeni v skupni rodovni knjigi kot ena pasma z dvema podtipoma, FCI pa jih je uradno priznala kot pasmo leta 1939. Pasm sta bili mednarodno ločeni v letu 1968, vendar pa je do takrat velikost populacije kraškega ovčarja že nevarno padla, zaradi česar so v pasmo vkrizali tudi dva novofundlandca [13]. Čeprav je tudi šarplaninec po

obeh svetovnih vojnah postal redkeje zastopan, se je vse do današnjih časov ohranila tudi lokalna populacija pastirskih psov, ki jih redno vključujejo v vzrejo pasme. Prav tako so v šarplaninca domnevno vkrižali kavkaške ovčarje. Kljub relativno nedavni uradni ločitvi v posamezne pasme odkrivamo v tej raziskavi presenetljivo raznoliko populacijsko dinamiko pri pasmah kraškega ovčarja, šarplaninca in tornjaka. Naše delo razkriva velik pomen genskega zdrsa (drifta) in omejenega genskega toka za populacijsko strukturo ter gensko razlikovanje pastirskih pasem Zahodnega Balkana.

## *2.2 Metode in rezultati*

### *2.2.1 Vzorci in izolacija DNA*

Zbrali smo vzorce krvi in dlake 493 psov treh pastirskih pasem psov Zahodnega Balkana (kraški ovčar, šarplaninec, tornjak) in dveh dodatnih pasem (kavkaški ovčar, novofundlandec), ki sta bili vkrižani v nekatere od preučevanih pasem Zahodnega Balkana. Vzorce smo pridobili preko pasjih razstav, pasemskih klubov, lastnikov in vzrediteljev, sočasno z njihovim soglasjem za uporabo vzorcev za namene populacijsko-genetske raziskave. Vzorci so bili zbrani v skladu z evropskimi smernicami za uporabo živali v raziskavah, krvne vzorce pa so odvzeli veterinarji v skladu z Zakonom o zaščiti živali. Raziskavo je odobril odbor za nego in uporabo živali v znanosti in raziskavah na Oddelku za zootehniko Biotehniške fakultete Univerze v Ljubljani. Kadar so bili na voljo rodovniški podatki, smo se izogibali vključevanju sorodnih živali. V primeru pasme kraškega ovčarja smo v analizi uporabili dve tretjini podatkov zbirke DNA, ki se z namenom preverjanja starševstva vzdržuje na Oddelku za zootehniko Biotehniške fakultete. DNA smo izolirali iz polne krvi in mešičkov dlake. Genomsko DNA iz zamrznjene krvi smo izolirali z uporabo komercialnega kita Gentra Puregene Blood kit (Qiagen). Za pridobivanje DNA iz dlake smo odrezali mešičke od pet do petnajstih dlak posameznega psa približno pet milimetrov od konca korenine in jih obdelali v pufru za lizo (1 M Tris-HCl (pH 8,3), 2 M MgCl<sub>2</sub>, 1 M KCl, 0,5% Tween 20) z 1,5 ul proteinaze K (20 mg/ml). Vzorce smo hranili preko noči pri 54°C in nato še deset minut pri 94°C.

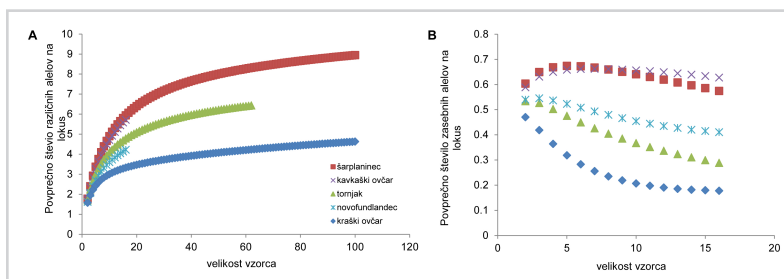
### *2.2.2 Genska pestrost na osnovi mikrosatelitnih označevalcev*

Osemnajst mikrosatelitnih označevalcev, ki jih priporoča Mednarodno združenje za živalsko genetiko (International Society for Animal Genetics, ISAG), smo pomnožili

na kapilarnem sekvenatorju ABI3130xl z uporabo komercialnega kita (Canine Genotypes™ Panel 1.1, Finnzymes). Za analizo genske pestrosti smo uporabili opaženo ( $H_O$ ) in pričakovano ( $H_E$ ) heterozigotnost ter koeficient molekularnega sorodstva (angl. coancestry,  $f$ ) v programu Molkin v 3.0 [14]. Ker so bile velikosti vzorcev različne, smo v tej analizi uporabili naključno podmnožico 30 osebkov vsake pasme. Število alelov na lokus (angl. allelic richness, AR) in zasebno število alelov na lokus (angl. private allelic richness, pAR) smo ocenili s postopkom razredčitve (angl. rarefaction) v programskem paketu ADZE [15] (slika 2.1).

Slika 2.1

Število alelov na lokus. (A) Povprečno število različnih in (B) zasebnih alelov na lokus kot funkcija standardizirane velikosti vzorca petih preiskovanih pasem.



Parametri pasemsko-specifične genske pestrosti so prikazani spodaj (tabela 2.1). Analiza molekulske variance je pokazala, da se 72% raznolikosti nahaja znotraj pasem. Kraški ovčar izkazuje najnižjo heterozigotnost (0,5997), medtem ko imajo druge pasme višje ocene heterozigotnosti. Najvišjo sorodnost, kot jo določa koeficient molekularnega sorodstva, smo ugotovili pri kraškem ovčarju (0,3936) in najnižjo pri šarplanincu (0,2098). Enako težnjo smo opazili tudi z razredčenimi števili alelov na lokus: najvišje število alelov na lokus smo odkrili pri šarplanincu (5,9429) in najnižje pri kraškem ovčarju (3,3268). Šarplaninec in kavkaški ovčar imata najvišje število zasebnih alelov.

Tabela 2.1: Povzetek genske pestrosti pasem glede na mikrosatelitne podatke.

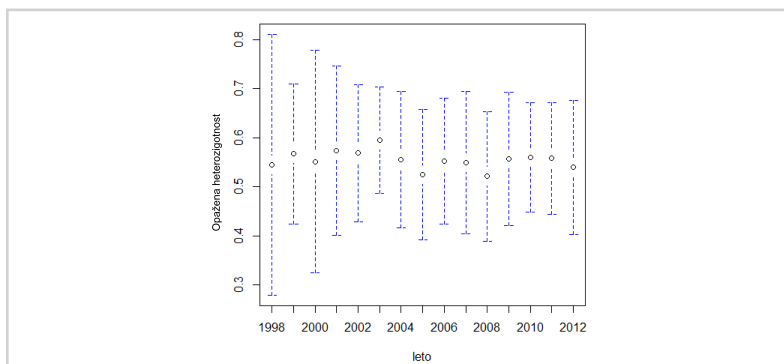
Pasma	N	$H_O$	$H_E$	AR	$f$
Kraški ovčar	326	0,5997	0,5775	3,3268	0,3936
Šarplaninec	109	0,7579	0,7703	5,9429	0,2098
Kavkaški ovčar	10	0,7661	0,6916	4,7839	0,2701
Tornjak	36	0,7603	0,7062	5,7164	0,2666
Novofundlandec	13	0,6013	0,5909	4,2183	0,3757

S programom Molkin v 3.0 smo določili tudi prispevke vseh treh pasem k pasemski, medpasemski in celokupni pestrosti glede na metodi Caballera in Tora [16] in Petita ter sodelavcev [17] (tabela 2.2). Izguba šarplaninca bi predstavljala največje zmanjšanje celotne genske pestrosti. Nasprotno pa bi zaradi nizke pasemske genske pestrosti kraškega ovčarja odstranitve te pasme imela za posledico relativen doprinos k celotni pestrosti. Izguba tornjaka bi povzročila zmanjšanje celotne pestrosti, in sicer v večji meri na račun medpasemske kot pa znotrajpasemske raznolikosti pasme.

Tabela 2.2: Prispevki pasem k genski pestrosti in številu alelov na lokus. Porast ali izgubo pestrosti smo izračunali z metodama Caballera & Tora [16] in prispevke k številu alelov na lokus z metodo Petita in sodelavcev [17].  $GD_W$ : porast ali izguba genske pestrosti znotraj populacij, če se pasma odstrani.  $GD_B$ : porast ali izguba genske pestrosti med populacijami, če se pasma odstrani.  $GD_T$ : porast ali izguba celotne genske pestrosti, če se pasma odstrani.  $AR_W$ : prispevek vsake populacije k številu alelov na lokus znotraj populacije.  $AR_B$ : prispevek vsake populacije k številu alelov na lokus med populacijami.  $AR_T$ : prispevek vsake populacije k celotnemu številu alelov na lokus.

Pasma	$GD_W$	$GD_B$	$GD_T$	$AR_W$	$AR_B$	$AR_T$
Kraški ovčar	8,080	-5,772	2,308	-15,686	9,930	-5,756
Šarplaninec	-7,026	-0,435	-7,460	14,491	5,178	19,669
Tornjak	-0,885	-1,855	-2,741	1,195	1,904	3,099

V časovnem okvirju, ki ga dopušča zbran genski material kraškega ovčarja, ni opazno izrazito spreminjanje heterozigotnosti pri tej pasmi (slika 2.2).



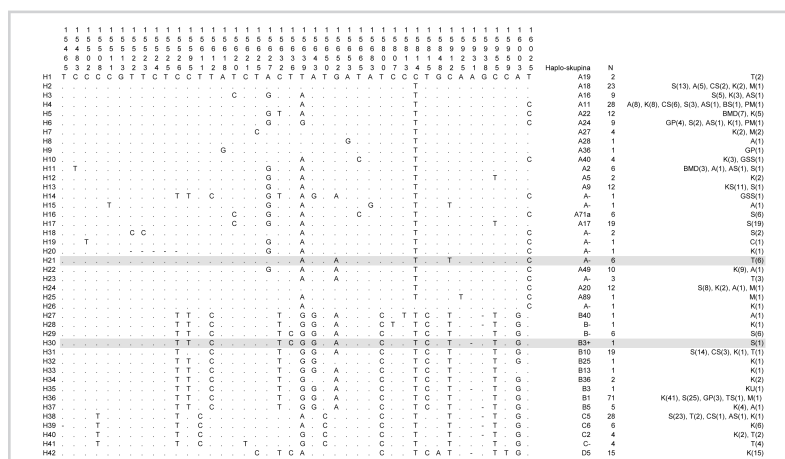
Slika 2.2

Opažena heterozigotnost glede na leto rojstva kraških ovčarjev.

### 2.2.3 Genska pestrost na osnovi mitohondrijske DNA

Z začetnimi oligonukleotidi in pogoji, opisanimi v [18], smo v 94 vzorcih pomnožili območje mitohondrijske D-zanke dolžine 680 bp. Podatke zaporedij mitohondrijskih D-zank pastirskih psov Zahodnega Balkana smo dopolnili z 231 zaporedij še 14 pastirskih pasem psov iz podatkovne zbirke GenBank. Zaporedja smo skrajšali na nukleotidne položaje 15458 – 16030 in s tem omogočili poravnave vseh objavljenih zaporedij. Vse mitohondrijske haplotipe smo odkrili s primerjavo z referenčnim zaporedjem kot opisano v [19]. Odkrili smo 42 različnih haplotipov. V sekvenciranih vzorcih smo odkrili dva nova haplotipa in ju vložili v podatkovno zbirko GenBank pod dostopnima števkama KF241537 in KF241538. Haplotip H30 (B3+, KF241538), ki se razlikuje v enem nukleotidu od haplotipa B3 (15636T->C), smo odkrili pri osebku šarplaninca, haplotip H21 (KF241537), ki se razlikuje v enem nukleotidu od haplotipa H23 (1912C->T), pa smo odkrili pri šestih psih pasme tornjak. Kar 86% raznolikosti mitohondrijske DNA (mtDNA) je znotraj pasem ( $\theta_{ST} = 0,141$ ,  $P < 0,0001$ ). Šestnajst (38%) haplotipov si je delilo več osebkov različnih pastirskih pasem (slika 2.3).

Za pasme z vsaj osmimi zaporedji smo s programom Arlequin v 3.5 [20] določili nukleotidno pestrost in pestrost haplotipov. Število haplotipov ( $k$ ) znotraj pasem se je gibalo od enega haplotipa do triindvajset haplotipov in pestrost haplotipov je bila od nič pri kraškem ovčarju, pri katerem smo odkrili le en haplotip (A9), do  $0,883 \pm 0,011$  pri šarplanincu (tabela 2.3).



Slika 2.3

Haplotype 42 D-zank. Prikazani so deleži haplotipov in porazdelitev po pasmah. Nukleotidni položaji so oštevilčeni kot v [19]. Istovetnost z referenčnim zaporedjem (A19) je prikazana s piko, substitucija s črko baze in deležija s pomišljajem. Nova haplotipa sta obarvana sivo. Okrajšave pasem so enake kot v mreži haplotipov, izrisani po postopku združevanja median (angl. median-joining). (slika 2.4).

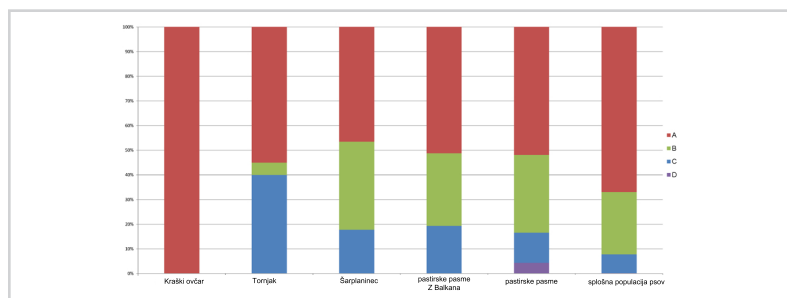
Tabela 2.3: Pestrost haplotipov in nukleotidna pestrost.

Pasma	N	k	Pestrost haplotipov	Nukleotidna pestrost
Kraši ovčar	11	1	0,000 ± 0,000	0,000 ± 0,000
Šarplaninec	129	15	0,883 ± 0,011	0,014 ± 0,007
Kavkaški ovčar	12	4	0,712 ± 0,105	0,011 ± 0,006
Tornjak	20	7	0,858 ± 0,047	0,012 ± 0,007
Kangal	114	23	0,840 ± 0,028	0,015 ± 0,008
Akbaš	20	9	0,800 ± 0,073	0,008 ± 0,004
Bernski planšarski pes	10	2	0,467 ± 0,132	0,002 ± 0,002
Pirenejski planinski pes	8	3	0,679 ± 0,122	0,013 ± 0,008

Porazdelitev haplotipov med pasmami smo določili z algoritmom združevanja median (angl. median-joining) v programu NETWORK v 4.6.1.1 (Fluxus Technology Ltd)<sup>1</sup>. Najbolj pogost haplotip (B1) smo opazili pri 71 osebkih petih različnih pasem, ki sta mu sledila dva druga haplotipa (A11 in C5), oba prisotna pri 28 osebkih. Mreža haplotipov (slika 2.4) kaže, da vsi pastirski psi Zahodnega Balkana pripadajo haplo-skupinam A, B in C.

<sup>1</sup>www.fluxus-engineering.com





Slika 2.5

Porazdelitev haplo-skupin. Porazdelitev haplo-skupin mtDNA je prikazana za posamezne pastirske pasme Zahodnega Balkana, skupaj s povprečnimi deleži haplo-skupin v celotnem vzorcu pastirskih psov Zahodnega Balkana, povprečnimi deleži preko vseh 16 pastirskih pasem in s povprečnimi deleži v splošni populaciji psov kot navedeni v [21].

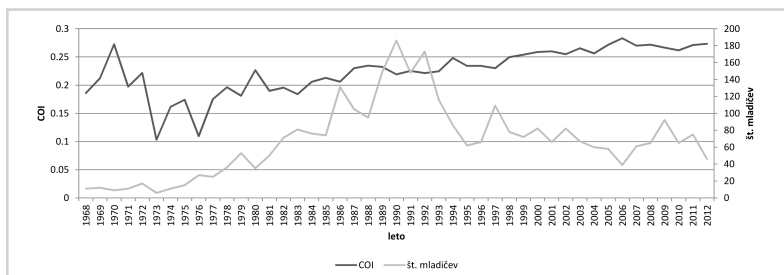
#### 2.2.4 Analiza rodovnikov

Rodovnika pasme kraškega ovčarja in tornjaka smo analizirali z uporabo programov ENDOG v 4.8 [22] in PEDIG [23]. Metode so podrobneje opisane v literaturi, na primer v [24]. Opredelili smo referenčni populaciji psov, rojenih med letoma 2000 in 2012. Za kraškega ovčarja je znašala ta 860 živali, za tornjaka pa 3534 živali. Rodovniške podatke kraškega ovčarja smo pridobili s pomočjo Društva ljubiteljev in vzrediteljev kraških ovčarjev (DLVKOS) in so javno dostopni v informacijskem sistemu<sup>2</sup>, ki ga društvo vzdržuje v sodelovanju z Oddelkom za zootehniko Biotehniške fakultete. Vključeni so skoraj vsi psi, ki jih je registrirala bivša Jugoslovanska kinološka zveza in kasneje Kinološka zveza Slovenije. Rodovniške podatke tornjaka smo pridobili iz spletne podatkovne zbirke rodovnikov<sup>3</sup> te pasme, ki predvidoma vključuje približno tri četrtine psov, registriranih na Hrvaškem in tretjino psov, registriranih v Bosni in Hercegovini (G. Visković, osebna komunikacija). Trenutna populacija kraševcev temelji na 19 ustanoviteljih, katerih poreklo ni znano, in dveh psov z enim znanim staršem. Trenutna populacija tornjaka ima 218 ustanoviteljev in 14 psov z enim znanim staršem. Povprečna globina rodovnika znaša za kraškega ovčarja 12,8 povprečnih enakovrednih generacij in le 4,2 povprečne enakovredne generacije za tornjaka, kar je skladno z nedavnim uradnim priznanjem te pasme. Vendar pa nizka globina rodovnika pomeni, da populacijski parametri tornjaka niso nujno primerljivi parametrom kraševca. Efektivna velikost ustanoviteljev kraškega ovčarja je 8,8, povprečen koeficient sorodstva 0,29, povprečen koeficient inbreedinga pa 0,27 in še narašča (slika 2.6). Tako se kraševcevr uvršča v družbo pasem z najvišjo dokumentirano stopnjo inbreedin-

<sup>2</sup><http://www.dlvkos.si/>

<sup>3</sup><http://tornjak.pedigree.net/>

ga. Povprečen koeficient inbreedinga irskega setra na osnovi 10 generacij znaša 16% [25], povprečen koeficient inbreedinga novoškotskega prinašalca na osnovi 13 generacij znaša 26% [24], povprečen koeficient inbreedinga poljskega goniča pa kar 37% [26]. Zaradi relativne narave koeficienta inbreedinga se pogosto raje uporablja stopnja inbreedinga ( $\Delta F$ ) in realizirana efektivna velikost populacije ( $1/2\Delta F$ ). V navedenem časovnem obdobju ocenjujemo realizirano efektivno velikost populacije kraškega ovčarja na 19,3 živali. Za primerjavo, realizirana efektivna velikost populacije tornjaka je 98,3 osebkov. V splošnem znaša priporočena minimalna efektivna velikost populacije pasme 100 živali. Efektivna velikost populacije pod 100 živali je v smislu ohranjanja pasme zaskrbljujoča. Če pa je nižja od 50 osebkov, velja na dolgi rok za nevzdržno. Najvišja stopnja inbreedinga, ki še omogoča nadomestitev izgubljene genske pestrosti, je 0,5 do 1 odstotna [27]. Pri kraševcu znaša 2,6% in kaže na to, da pasma izgublja gensko pestrost hitreje, kot jo lahko nadoknadi. Prav tako so genski prispevki ustanoviteljev pasme kraškega ovčarja strogo neenakomerni. Deset ustanoviteljev prispeva 87% genskega sklada v referenčni populaciji. S samo tremi ustanovitelji lahko razložimo kar 52% genske pestrosti v referenčni populaciji in s samo enim ustanoviteljem kar 36% genskega sklada. Tabela 2.4 prikazuje genske prispevke vseh psov ustanoviteljev.



Slika 2.6

Koeficient inbreedinga (COI) ter število mladičev kraškega ovčarja skozi čas.

Glede na pasmo ustanoviteljev izvira kar 43% genskega sklada iz novofundlandcev, 26% iz psov, uvoženih iz Šarplanine, 24% iz lokalnih psov Krasa in 7% iz psov negotovega porekla. V trenutni populaciji obstaja samo ena materinska linija, kar je skladno z odkritim enim haplotipom mtDNA v tej pasmi. Od leta 2008 ima skoraj celotna populacija kraškega ovčarja rutinsko potrjene rodovnike s testom starševstva. To pomeni, da je možnost za napake v zadnjih letih močno znižana. Napako rodovnikov smo ocenili s pristopom ABC (angl. approximate Bayesian computation) [28] na povprečno

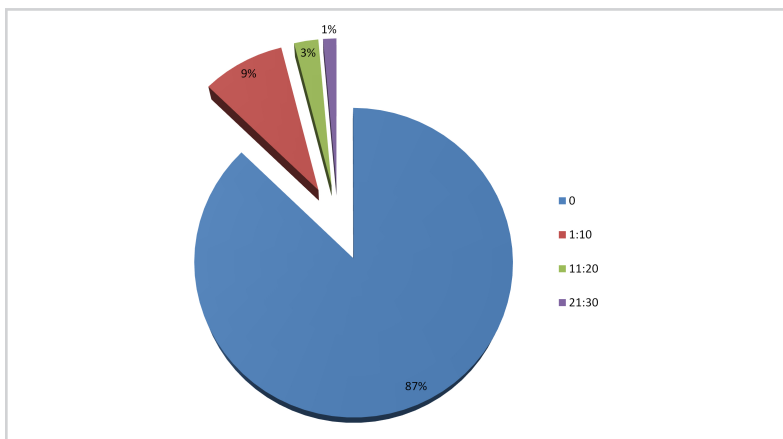
Tabela 2.4: Genski prispevki ustanoviteljev populacije kraških ovčarjev, rojenih med letoma 2000 in 2012.

Ustanovitelj	Spol	Leto rojstva	Prispevek ustanovitelja	Pasma / Izvor	Prispevek glede na izvor ustanovitelja		
Brk I	M		0,357	Novofundlandca	0,428		
Brk II	M	1973	0,071				
Karamashi I	M	1926	0,054	psi s Šarplanine	0,263		
Bilić	M	1945	0,049				
Sarija	F	1945	0,049				
Bistra	F	1955	0,038				
Kandil	F	1933	0,019				
Lisič Mirnski	M	1937	0,018				
Fafila	F		0,011				
Karabaš Gare	M		0,011				
Vidra	F	1929	0,009				
Karamashi II	M	1927	0,003				
Cura	F	1929	0,002				
Loro	M	1946	0,089			lokalni psi s Krasa	0,241
Živa	F	1952	0,062				
Baba	F	1946	0,045				
Sama	F		0,022				
Pikolo	F	1920	0,012				
Marko	M	1920	0,012				
Žarko	M	1976	0,057	negotov izvor	0,067		
Siva	F	1946	0,010				

vrednost 6,6% (interval zaupanja 0–18%), kar nakazuje prisotnost napak v rodovnikih v generacijah pred uvedbo testiranja starševstva. Velikost in razpon napake sta podobna ocenam napak pri drugih pasmah psov [28].

Čeprav v celotnem rodovniku pasme odkrijemo le tri parjenja med brati in sestrami (0,09% vseh paritev), 25 parjenj med polbrati in polsestrami (0,77% vseh paritev) ter

12 paritev med staršem in njegovim mladičem (0,37% vseh paritev), podatki kažejo na neenakomerno uporabo psov v reprodukciji tudi po letu 2000 (slika 2.7). Velika večina psov ni vključena v vzrejo (87%), ali pa je imela manj kot 10 mladičev (9%). Po drugi strani je 12 psov (1%; 7 samcev, 5 samic) imelo več kot 20 mladičev. Največje število potomcev posameznega plemenskega psa oz. psice, rojene(ga) po letu 2000, je 30. Povprečno število potomcev na plemenjaka je 9,7, na plemenko pa 9,5 mladičev.

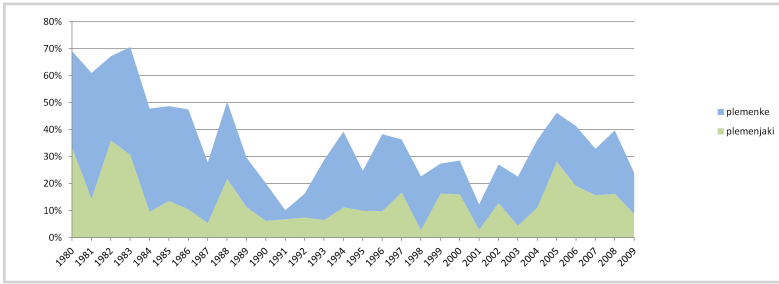


Slika 2.7

Porazdelitev živali glede na število potomcev. Kar 87% psov, rojenih po letu 2000, nima potomcev.

Delež živali celotne populacije, ki je bil vključen v vzrejo, je bil za samce nižji kot za samice. Mednarodna kinološka zveza (FCI) v splošnem priporoča, da se iz rejskega programa ne izključi več kot 50% pasme. Od kraševcev, rojenih po letu 2000, je bilo v povprečju v vzrejo vključenih 18% samic. Pri samcih je ta delež 13%. Kasneje sta se deleža za oba spola nekoliko povečala. Na primer, od psov, rojenih leta 2008, je bilo do danes vključenih 24% samic in 16% samcev (slika 2.8). V nasprotju pa je bil v letih osamosvojitve Slovenije, kljub velikemu številu rojstev psov, delež psov, rojenih v tem času, ki so bili vključeni v vzrejo, relativno nizek. Kot primer, od živali, rojenih leta 1991, je bilo v vzrejo vključenih le 7% samcev in le 3% samic.

Posledica parjenja v sorodstvu se običajno najprej izrazi v pojavu depresije zaradi inbreedinga (inbreeding depresija). Povprečna velikost legel kraških ovčarjev od leta 2000 naprej znaša 5,3 mladičev ( $SD=0,52$ ), podatki o velikostih legel pred tem letom pa so manj zanesljivi (npr. registracije samo 6 mladičev v leglu). Prav tako je nizko štiriletno povprečje uspešnosti paritev, saj znaša 56,6%. Neredko se pri vzreji kraševcev



Slika 2.8

Delež živali, vključenih v vzrejo, glede na leto rojstva.

uporablja tudi umetno oploditev. Za ugotavljanje trendov vpliva genske pestrosti na fitnes pasme bi bilo potrebno tovrstne podatke zasledovati v daljšem časovnem okviru, kot je bilo mogoče v tej raziskavi.

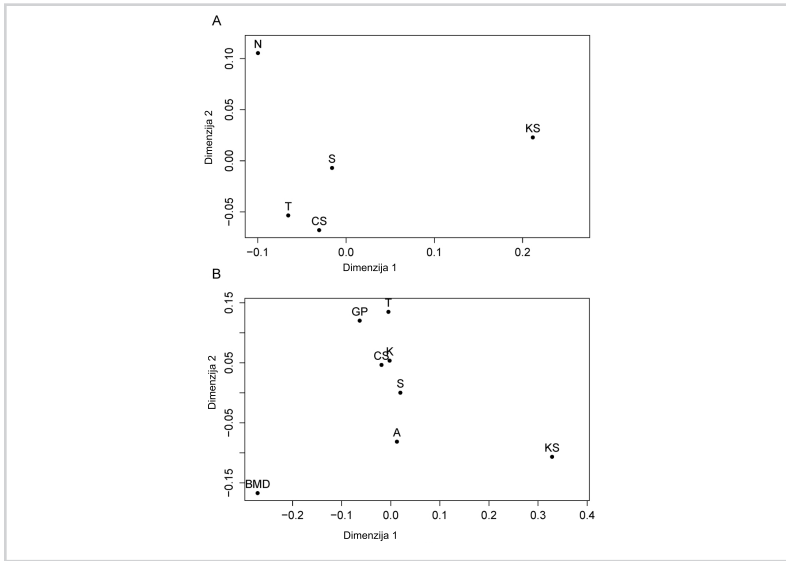
### 2.2.5 Razlikovanje pasem

Gensko razlikovanje med pasmami smo analizirali glede na vrednosti koeficientov  $F_{ST}$  (10 000 permutacij) v programu Arlequin v 3.5 [20]. Za predstavitev razdalj med pari pasem smo uporabili večdimenzijsko lestvičenje (angl. multidimensional scaling, MDS) (slika 2.9). Ocene genskega razlikovanja so vse značilne ( $P < 0,05$ ), kar potrjuje, da lahko preučevane pasme veljajo za kot gensko neodvisne. Vrednosti  $F_{ST}$  razkrivajo najnižje razlikovanje med šarplanincem in kavkaškim ovčarjem. Kraški ovčar in novofundlandec sta bolj oddaljena od drugih, preostale pasme pa se združujejo bližje skupaj. Povprečne vrednosti  $F_{ST}$  so v negativni soodvisnosti z opaženimi heterozigotnostmi pasem ( $r = -0,923$ ). Prav tako smo analizirali koeficiente  $F_{ST}$  na ravni mtDNA (slika 2.9 B). Pasm z najnižjim številom haplotipov (kraški ovčar in bernski planšarski pes) sta v prikazu MDS umeščeni ločeno, druge pasme pa imajo vmesni položaj in so medsebojno bližje povezane.

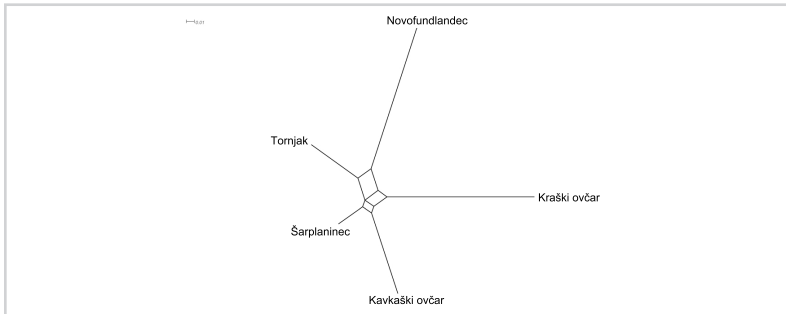
Nei-jeve razdalje ( $D_A$ ) med pasmami smo izračunali s programom POPULATIONS v 1.2.32 [29]. Za izris filogenetskega drevesa pasem s pomočjo razdalj  $D_A$  smo uporabili metodo Neighbor-Net [30] s programom SplitsTree4 v 4.12.6 [31]. Rezultati te analize podpirajo ocene genskega razlikovanja. Kraški ovčar se loči od središča kot neodvisna in bolj oddaljena veja, šarplaninec in kavkaški ovčar pa kažeta medsebojno bližino (slika 2.10).

*Slika 2.9*

Analiza večdimenzijskega lestvičenja (MDS), ki smo ga izvedli na osnovi vrednosti (A) mikrosatelitnih  $F_{ST}$  in (B) mtDNA  $F_{ST}$ . Kratice pasem so enake kot v sliki mreže haplotipov (slika 2.4).

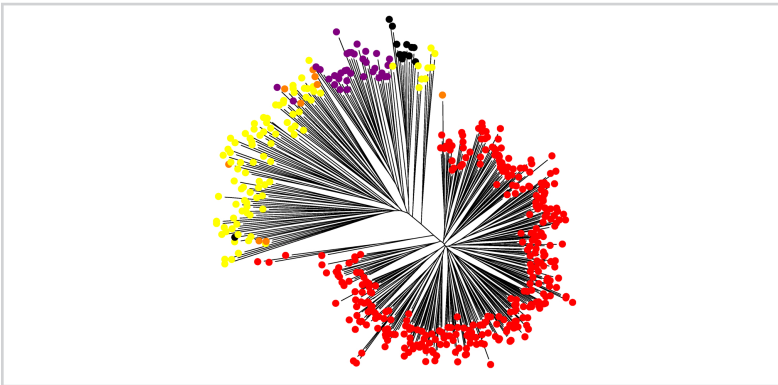
*Slika 2.10*

Drevo analize Neighbor-Net za pet preučevanih pasem.



### 2.2.6 Populacijska struktura

Za predstavitev odnosov med osebki smo uporabili metodo združevanja sosedov (angl. neighbor-joining, NJ) s programom PHYLIP [32] in uporabo razdalj med aleli na lokusu (angl. DAS, shared allele distances). Razdalje DAS med osebki smo izračunali s programom POPULATIONS v 1.2.32 [29] in jih predstavili v obliki drevesa (dendrograma). Večina živali posamezne pasme je blizu skupaj, vendar so prisotne tudi izjeme (slika 2.11). Topologijo drevesa jasno določata dve skupini: kraški ovčarji in preostale pasme. Kavkaški ovčarji ne oblikujejo lastne veje, ampak so razpršeni v vejah šarplaninca. En osebek je v skupini kraških ovčarjev. Tudi nekaj živali drugih dveh pasem je napačno umeščenih v skupino šarplanincev. Najverjetnejši razlog za to je visoka genska pestrost te pasme ter manjše število označevalcev v analizi. Nekaj šarplanincev tvori ločene veje, ki so nekoliko odmaknjene od glavne skupine šarplanincev, kar namiguje na možnost populacijske razslojenosti vzorcev šarplanincev.

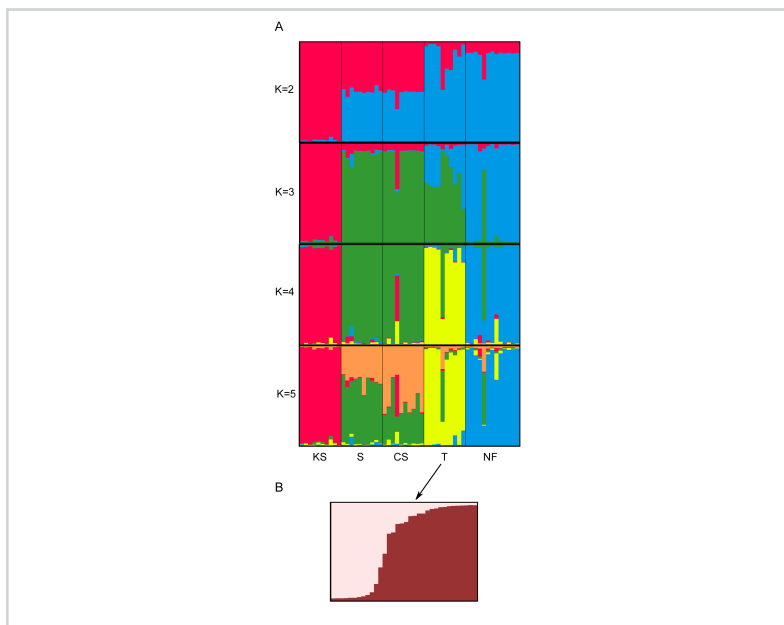


Slika 2.11

Drevo združevanja sosedov, izrisano na osnovi razdalj DAS med vsemi živalmi. Vsaka barvna pika predstavlja posamezno žival. Pasma pripadajo sledečim barvam: rumeno, šarplaninec; rdeče, kraški ovčar; vijolično, tornjak; oranžno, kavkaški ovčar; črno, novofundlandec.

Nenadzorovano razvrščanje v genetsko-populacijske skupine smo ugotavljali s programom Structure v 2.3.4 [33]. Simulacije smo izvedli brez vhodnih podatkov o pasemski pripadnosti živali z modelom, ki predpostavlja križanje in soodvisne alelna pogostosti v različnih populacijah, s 50 000 začetnimi koraki, izločenimi iz analize (angl. burn-in) in 100 000 koraki MCMC (Monte Carlo Markove verige). Za vsako domnevno število skupin ( $K$ ) smo opravili deset neodvisnih simulacij s  $K$  od dva do pet. Najbolj verjetno število skupin v našem vzorcu smo določili za  $K=4$  z metodo Evanna in sodelavcev [34]. Pri tem  $K$  se ločijo vse pasme razen šarplaninca in kavkaškega ov-

čarja (slika 2.12). Kraški ovčar je odstopil od preostalih pasem že po drugem izračunu skupin ( $K=3$ ). Nasprotno pa algoritem ni uspel razlikovati med šarplanincem in kavkaškim ovčarjem niti pri  $K=5$ . Preostale pasme so ostale v ločenih skupinah z ocenami pripadnosti  $>85\%$ . Ko smo izločili šarplaninca in kavkaškega ovčarja iz celotnega nabora podatkov ter ju analizirali ločeno pri  $K=2$ , se pasmi še vedno nista razdelili v dve skupini. Analizo smo ponovili s podmnožicami podatkov, da bi ocenili populacijsko strukturo posameznih pasem. Odkrili smo razslojenost v vzorcih tornjaka (slika 2.12 B), kar bi lahko bila posledica različne predniške sestave dveh skupin. Tornjak je tudi edina pasma z značilnimi odkloni od Hardy-Weinbergovega ravnotežja ( $P < 0,005$  z Bonferronijevim popravkom). Heterozigotnost je v tej pasmi višja od pričakovane, kar bi lahko bila posledica križanja predhodno ločenih populacij, na primer hrvaške in bosanske populacije pasme.



*Slika 2.12*

Bayesovo združevanje skupin pastirskih pasem psov Zahodnega Balkana in dveh dodatnih pasem, pomembnih v zvezi z vkrižanjem. (A) Analiza s programom Structure za vrednosti K od dva do pet. (B) Ločena analiza s programom Structure za pasmo tornjak kaže dve skupini. Kratice pasem so enake kot v mreži haplotipov (slika 2.4).

Verjetnost, da bo osebek dodeljen v eno izmed petih pasem, smo ocenili s koeficientom pripadnosti pasmi (Q) v programu Structure in s pristopom [35] programa

GeneClass2 [36]. Z dodeljevanjem v pasme [35] smo 98,6% (489 od 496) psov dodelili v pravilno pasmo. Deleži pravilne dodelitve so bili v razponu od 90% (kavkaški ovčar) do 100% (kraški ovčar) (tabela 2.5). Kljub temu so trije kavkaški ovčarji in dva tornjaka (sicer pravilno dodeljeni) izkazovali >10% posteriorno verjetnost pripadanja pasmi šarplaninca in šest šarplanincev pasmi kavkaškega ovčarja.

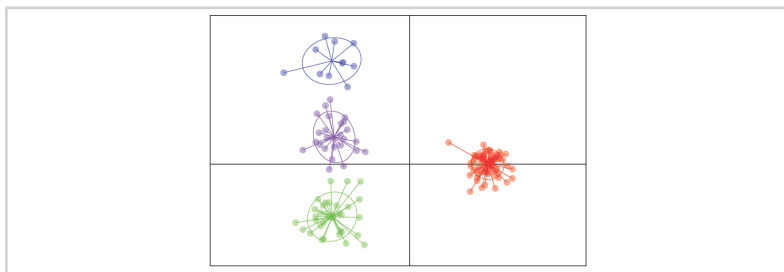
Tabela 2.5: Pripadnost pasemski skupini in dodeljevanje pasmam.

Pasma	Pripadnost pasemski skupini (Q) (Structure)	Dodeljevanje s pristopom Rannala & Moun-tain (1997) [35] (Gene-class2) (%)
Kraški ovčar	0,9523	100
Šarplaninec	0,6078	99,7
Kavkaški ovčar	0,5535	90,5
Tornjak	0,8482	99,9
Novofundlandec	0,8708	93,4

Opravili smo tudi diskriminantno analizo glavnih komponent (angl. discriminant analysis of principal components, DAPC) z uporabo paketa ADEGENET [37]. Razvrščanje z zaporednimi k-povprečji (angl. k-means) smo izvedli z Bayesovim informacijskim kriterijem (BIC) in tako določili ustrezno število skupin. Nato smo opravili diskriminantno analizo glavnih komponent za  $K=4$  in ohranili 40 komponent analize glavnih komponent (angl. principal component analysis, PCA) (87,9% celotne variance) in tri osi diskriminantne analize. DAPC analiza, ki ni odvisna od predpostavke o Hardy-Weinbergovem ravnotežju, je potrdila razvrstitev pasem v skupine s programom Structure. Pasma in skupine so se skoraj popolnoma ujemale, razen pri šarplanincu in kavkašem ovčarju, ki sta pripadala eni skupini in se ju ni dalo razločiti (slika 2.13). Vodoravna os DAPC analize je ločila kraškega ovčarja od preostalih pasem, navpična os pa je razdelila preostale tri skupine.

*Slika 2.13*

Diskriminantna analiza glavnih komponent (DAPC). Na sliki je prikazan graf prvih dveh glavnih komponent DAPC. Analiza DAPC je razkrila 4 skupine. Modra, novofundlandec; vijolična, šarplaninec in kavkaški ovčar; zelena, tornjak; rdeča, kraški ovčar.

*2.3 Diskusija in prihodnji izzivi*

V tem poglavju smo z namenom boljšega razumevanja strukture populacij in osnovnih genskih mehanizmov, ki vplivajo na razločevanje sorodstvenih odnosov med pasmami, z uporabo različnih pristopov analizirali podatke mikrosatelitnih označevalcev, zaporedij mtDNA in rodovniških podatkov. Pastirske pasme Zahodnega Balkana so primer učinkov relativno sodobne spremembe v načinu vzreje psov. Navkljub verjetnemu skupnemu izvoru pasem smo odkrili velike razlike v genski pestrosti teh sorodnih pasem. Glede na podatke mikrosatelitov in mtDNA izkazuje kraški ovčar nižjo gensko pestrost in se ostro razlikuje od drugih pasem. Izločitev kraškega ovčarja iz nabora podatkov povzroči relativno povečanje celokupne pestrosti, kar še dodatno predstavlja to pasmo kot gensko neraznoliko v primerjavi z drugimi preučevanimi pastirskimi pasmami Zahodnega Balkana. Kot kaže analiza rodovnikov kraških ovčarjev, je bil močan ustanoviteljev učinek (angl. founder effect) še ojačan z majhno efektivno velikostjo populacije in neenakomernimi prispevki živali v vzreji preko generacij, kar potrjuje, da se je genska erozija zgodila z genskim zdrsom. V izogib nadaljnji izgubi genske pestrosti bi bilo potrebno povečati število psov v vzreji, izenačiti njihove reprodukcijske doprinese ter omejiti število mladičev na vzrejno žival in tako povečati celokupno število plemenskih psov. Povprečno število potomcev na plemensko žival je 9 mladičev. Priporočljivo je, da posamezen pes v vzreji ne prispeva mnogo več potomcev od povprečja. Priporočljivo bi tudi bilo, da se omogoči raznolikost pasme z izogibanjem selekciji za fenotipske lastnosti, ki niso pomembne za zdravje in fitnes živali (npr. barva dlake in oči). Smiselno bi bilo razmisliti o vzpostavitvi rednega genskega toka v to pasmo. Čeprav je kraški ovčar izmenjeval svoj genski material s šarplanincem do leta 1968 in čeprav kar 43% genskih prispevkov izvira iz novofundlandcev, vpliv teh pasem na gen-

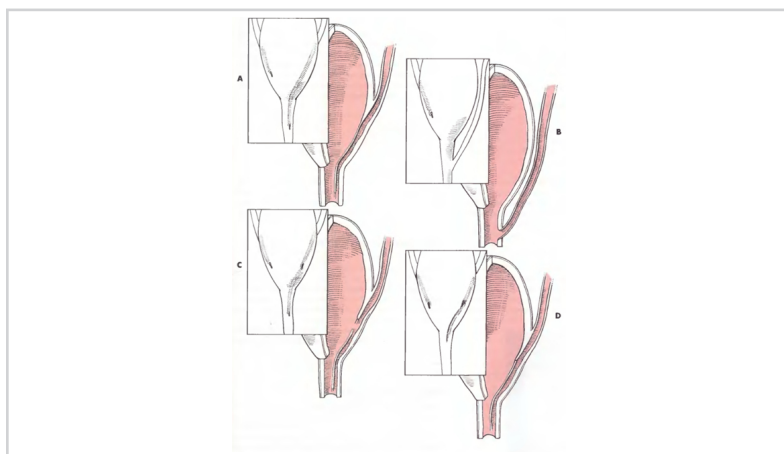
ski sklad kraškega ovčarja ni očiten iz trenutne populacijske strukture ali razlikovanja pasem. Razhajanje te pasme od preostalih potrjujejo ocene razlikovanja, položaj pasme v drevesu Neighbor-Net, razločitev z analizama Structure in DAPC, kot tudi ločenost kraških ovčarjev od osebkov vseh drugih pasem v drevesu združevanja sosedov. Prav tako smo v pasmi odkrili le en redek haplotip mtDNA (A9). Rezultati torej poudarjajo, da so za populacijsko strukturo redkih pasem večjega pomena stohastični kot pa deterministični demografski procesi. Na osnovi ocen razlikovanja so si šarplaninec, tornjak in kavkaški ovčar gensko bližji. Še posebej šibko je razlikovanje šarplaninca in kavkaškega ovčarja, kar podpirajo tudi analiza Structure, visoka stopnja navzkrižnih dodelitev in slaba ločitev pasem v drevesu združevanja sosedov. To je v skladu z že odkrito nizko gensko razdaljo med šarplanincem in kavkaškim ovčarjem [38]. Vendar pa je bilo število vzorcev kavkaškega ovčarja in novofundlandca v raziskavi nizko, saj sta bili pasmi vključeni v analizo z namenom preverjanja možnosti znavanja medpasemskega križanja. V kombinaciji z manjšim številom mikrosatelitnih označevalcev bi lahko bila tudi majhna velikost vzorcev odgovorna za težavno razlikovanje med šarplanincem in kavkaškim ovčarjem in nezaznan signal vkrižanja novofundlandcev v pasmo kraškega ovčarja. Vendar pa skladnost različnih analitskih pristopov kaže na relativno zanesljivost naših ocen. Šibko gensko razlikovanje je bilo odkrito tudi pri portugalskih pastirskih pasmah, ki jo avtorji razlagajo kot posledico dejstva, da ta tip psov ne predstavlja zaprtih populacij, saj se v pasmo še vedno vnaša tudi neregistrirane živali [3]. Podobna razlaga drži tudi za pastirske pasme Zahodnega Balkana, saj imata tako tornjak kot šarplaninec še vedno aktivni populaciji lokalno prilagojenih psov pastirskega tipa, ki se s pogojnimi rodovniki vključujejo v vzrejo registriranih pasem. Prekrivanje vseh haplotipov mtDNA kavkaškega ovčarja s haplotipi šarplanincev potrjuje, da je povezava med balkanskimi pasmami in kavkaškimi pastirskimi psi obstajala vsaj občasno v zgodovini, oziroma je rezultat nedavnega križanja. Šarplaninec in tornjak imata visoko gensko pestrost, raznoliko materinsko poreklo in novo odkrite haplotipe D-zanke mtDNA. Negativna soodvisnost med pasemsko gensko pestrostjo in razlikovanjem pasem je skladna z ugotovitvijo povezanosti pasemske heterozigotnosti in genske razdalje do drugih pasem [2]. Kljub temu so Huson in sodelavci [39] odkrili edinstven genski podpis (angl. genetic signature) celo v populaciji aljaških vlečnih psov križancev. Zato sklepamo, da spremembe alelnih pogostosti proti bolj ekstremnim vrednostim, ki povzročajo močno razhajanje pasem, niso posledica samo znižane heterozigotnosti, temveč v majhnih populacijah v večji meri tudi genskega zdrsa. V primeru križanj z

več različnimi pasmami, kot je to v primeru kraškega ovčarja, je takšno mešanje pasem še težavneje razločiti. Poleg genskega zdrsa so za razhajanja pasem lahko odgovorna tudi dodatna nedokumentirana križanja pasem, ki jih v tej raziskavi nismo preiskovali. Kadar so genski označevalci vezani z lokusi pod razdiralno selekcijo (angl. disruptive selection), lahko to dodatno napihne oceno genskega razlikovanja. Za preučevanje različnih procesov, ki vplivajo na porazdelitev genske pestrosti in odkrivanje njihovih prispevkov k strukturi populacij, bi bilo informativno analizirati tudi lokuse pod selekcijo. Dodatna razjasnitev izvora sil, ki vodijo k razhajanju populacij, je nujna, če želimo razumeti populacijske vzorce in njihove genske vplive. Za izboljšanje trenutnega znanja o demografski zgodovini, odnosih med populacijami psov in raznolikosti pasem bodo ključne prihodnje raziskave z uporabo visokozmogljivostnih genotipizacij, sodobnih metod sekvenciranja v kombinaciji z novimi statističnimi in računskimi metodami za odkrivanje in ugotavljanje preteklih dogodkov genskega mešanja [40–43].

*Odkrivanje genskega ozadja  
ektopije sečevodov pri kraškem  
ovčarju*

### 3.1 Ozadje

Ektopija sečevodov je najbolj pogost vzrok uhajanja seča pri mladih psih [44]. Razlog je nepravilen embrionalni razvoj ledvičnih (metanefričnih) in praledvičnih (mezonefričnih) vodov, pri čemer se ustje sečevoda pripenja kavdalno od trikotnika sečnega mehurja v sečni mehur (intravezikularno) ali v sečnici, prostati, semenovodu, nožnici ali maternici (ekstravezikularno). Glede na potek sečevoda v grobem razločimo ektopične sečevode, ki potekajo ob steni mehurja kavdalno od trikotnika mehurja (intramuralne), in sečevode, ki mehur popolnoma obidejo (ekstramuralne). Obstaja pa tudi več drugih oblik ektopije sečevodov (slika 3.1). Pogosto so prisotne tudi druge nepravilnosti sečnih organov, npr. razširjen sečevod zaradi zastajanja seča (hidroureter), razširjen ledvični meh in ledvične čašice zaradi zapore odtekanja seča (hidronefroza), gnojno vnetje ledvičnega meha (pielonefritis), zaostanek v razvoju mehurja (hipoplazija mehurja). Več ektopičnih sečevodov so opazili pri mnogih pasmah [45–49], vendar pa so vzročne genske spremembe še vedno večinoma nepoznane. Glede na spol in starost osebkov, mesto pripenjanja in potek sečevodov ter vpletenost enega ali obeh sečevodov so klinični znaki lahko raznoliki. V večcentrični presejalni raziskavi ektopije sečevodov v evropski populaciji pasme entlebuški planšarski pes je imelo le 36% psov ustja sečevodov v fiziološko pravilnem položaju, vendar pa je od 308 pregledanih psov kazalo znake uhajanja seča le 25 psov [48]. Podobne rezultate kaže nedavna raziskava 552 psov te pasme, v kateri so ugotovili, da ima pri 47,3% psov vsaj eden od sečevodov intravezikularno ektopično ustje, pri 19,8% psov pa vsaj eden od sečevodov ekstravezikularno ustje [50]. V primeru severnoameriške populacije te pasme so raziskovalci prav tako odkrili več primerov psov brez kliničnih znakov s potrjeno diagnozo ektopije sečevodov in kot temeljni mehanizem predlagali nizko stopnjo penetrance genov [47]. V zadnjih dveh desetletjih je bilo odkritih več primerov ektopije sečevodov pri pasmi kraški ovčar. Kot opisano v poglavju 2 znaša učinkovita velikost populacije kraškega ovčarja manj kot 20 živali, zaradi česar je pasma še toliko bolj dovzeta za genska obolenja. Po drugi strani predstavljajo ravno pasme psov z majhnim genskim skladom dober model za raziskave enogenskih (monogenih) in kompleksnih genskih bolezni, saj je vezavno neravnovesje (angl. linkage disequilibrium, LD) znotraj pasem veliko, haplotipi pa dolgi, kar omogoča asociacijske preiskave celotnega genoma (angl. genome wide association studies, GWAS) z manjšim številom SNP-jev in manjšim številom osebkov kot pri enakih raziskavah na ljudeh, kjer je genska raznovrstnost običajno mnogo večja.



Slika 3.1

Različni tipi ektopičnih sečevodov: (A) intramuralni, (B) ekstramuralni, (C) dvojne uretralne odprtine, (D) uretralna korita. Vir slike: [51].

## 3.2 Metode in rezultati

### 3.2.1 Izjava o etiki

Raziskavo smo opravili v skladu z narodnimi in mednarodnimi smernicami za dobro počutje živali v raziskavah. Delo je odobril odbor za nego in uporabo živali v znanosti in raziskavah na Oddelku za zootehniko Biotehniške fakultete Univerze v Ljubljani. Vsi psi, udeleženi v tej raziskavi, so last zasebnih lastnikov in so bili vključeni v raziskavo izključno z njihovo privolitvijo. Vse krvne vzorce so odvzele usposobljene osebe na veterinarskih klinikah, pridobili pa smo jih s pomočjo Društva ljubiteljev in vzrediteljev kraških ovčarjev (DLVKOS).

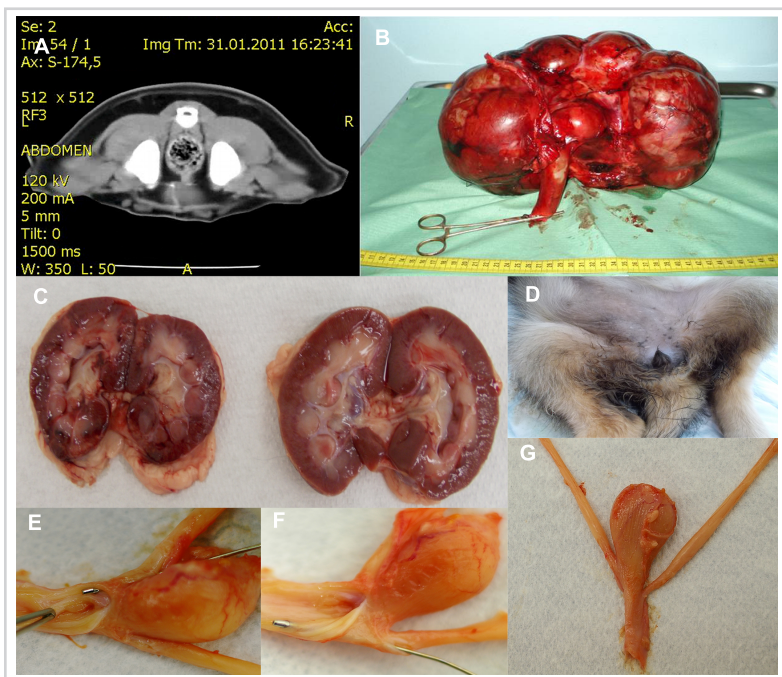
### 3.2.2 Fenotipski podatki in analiza rodovnikov

Fenotipske podatke in diagnostično gradivo smo zbirali s pomočjo lastnikov psov, vzrediteljev in veterinarjev v obdobju treh let (2010 - 2013). Mediana starosti ob diagnozi je pri dokumentiranih primerih kraškega ovčarja tri mesece, skladno z navedbami iz literature, da je ektopija sečevodov tipično ugotovljena pri mlajših psih [44]. Razmerje med spoloma med dokumentiranimi primeri ektopije sečevodov kraškega ovčarja je 5:2 ( $n=14$ ) z večjim deležem samic. Večina dokumentiranih primerov obišče veterinarske klinike s kliničnimi znaki uhajanja seča (slika 3.2 D). Spodaj (slika 3.2 A) je prikazan prečni posnetek vijačne računalniške tomografije (angl. computer tomography, CT)

kavdalnega sečnega mehurja prizadete dvomesečne samicke kraškega ovčarja s sumom diagnoze intravezikularnega ektopičnega sečevoda. Izločanje kontrastnega sredstva je vidno iz levega, normalno umeščenega sečevoda, a ni opazno iz desnega sečevoda, kar namiguje na to, da se desni sečevod odpira ektopično. Spodnji primer (slika 3.2 E-G) prikazuje desni ektopični sečevod pri dvomesečni psički z nezmožnostjo zadrževanja seča (inkontinenca), pri kateri je opravljena posmrtna preiskava trupla (nekropsija) potrdila kavdalni položaj ustja desnega sečevoda glede na levi sečevod.

### Slika 3.2

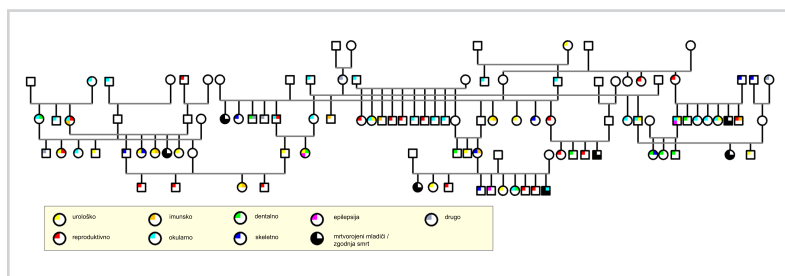
Ektopija sečevodov in nekatera druga obolenja sečil pri kraškem ovčarju. (A) Prečni posnetek vijačne računalniške tomografije primera desnega intravezikularnega ektopičnega sečevoda. Oba sečevoda se je dalo slediti do kraniolateralnega področja pred trikotnikom sečnega mehurja, vendar pa je bilo izločanje kontrastnega sredstva vidno samo iz levega sečevoda (slika uporabljena z dovoljenjem I. Firm). (B) Policistična ledvica pri sedemmesečnemu kraševcu (slika uporabljena z dovoljenjem C. Mezgec). (C) Hidronefroza in pielonefritis pri dvomesečni kraševki z ektopijo sečevodov. (D) Čep rav se ektopija sečevodov ne izrazi vedno kot uhajanje seča, pa je to najbolj pogost znak, zaradi katerega se privede prizadete živali na veterinarsko kliniko. (E) Normalen položaj ustja levega sečevoda in (F, G) nepravilen položaj ustja desnega sečevoda kavdalno od trikotnika sečnega mehurja (slike E-G uporabljene z dovoljenjem A. Domanjko-Petrič).



Poleg ektopičnih sečevodov so v pasmi pogoste tudi druge urološke bolezni, npr. okužbe sečil, pojav kamnov v sečilih (urolitiaz), policistična bolezen ledvic (slika 3.2 B), razširjen ledvični meh in ledvične čašice (hidronefroza) in gnojno vnetje ledvic (pielonefritis) (slika 3.2 C), razširjen sečevod (hidroureter), kronična okvara ledvic...

Pogosto so prisotni tudi le blagi klinični znaki, npr. pogosta žeja in uriniranje, ki pa z urogenitalnim razvojem mladih živali nemalokrat tudi prenehajo. Poleg nagnjenosti k boleznim sečil je pasma dovzeta za nekatere bolezni okostja (npr. hipertrofična osteodistrofija, kortikalna hiperostoza, displazija kolkov), zobovja (napake pri ugrizu, napačno število in položaj zob), oči (distihiaza, entropij, ektropij), spolnih organov (kriptorhidizem, rak na modih, rak prostate, distocija, mrtvorajeni mladiči), imunskega sistema (levkemije, avtoimuna obolenja kože, piodermija, alergije na hrano) in epilepsijo.

Rodovniške podatke pasme je zagotovilo Društvo ljubiteljev in vzrediteljev kraških ovčarjev (DLVKOS)<sup>1</sup>, dodatno pa njihovo pravilnost preverjamo z uporabo 19 mikrosatelitnih označevalcev, ki jih priporoča Mednarodno združenje za živalsko genetiko (International Society for Animal Genetics, ISAG) in smo jih uporabili tudi pri populacijski analizi kraškega ovčarja in sorodnih pasem (poglavje 2). Večgeneracijski rodovnik 111 psov z znanimi fenotipi (slika 3.3) smo pripravili z uporabo programa GenoPro<sup>2</sup>.



Slika 3.3

Primer rodovnika s 111 kraškimi ovčarji z obolenji sečil in drugimi predvidoma dednimi stanji. Sorodni psi iz legel z normalnim ali nepoznanim fenotipom so zaradi enostavnejšega izrisa izpuščeni. Simboli rodovnika so standardni za samca (kvadrat) in samico (krog). Definicije fenotipskih skupin so v legendi pod rodovnikom.

V rodovniku so prikazani psi z obolenji in psi, ki so imeli oz. imajo prizadete potomce glede na navedene fenotipske skupine bolezni. Sorodni psi z normalnim ali neznanim fenotipom so za namen enostavnejšega izrisa v rodovniku izpuščeni. Od 111 psov v analizi rodovnikov jih je 19 (13 samic, 6 samcev) imelo obolenje sečil. Takšno stanje lahko preskakuje generacije in prizadeti psi imajo večinoma starše brez kliničnih znakov, a je z uhajanjem seča pogosto prizadetih več mladičev v leglu. Obstajajo tudi manj zanesljivi in nepotrjeni dokazi, da naj bi bili z uhajanjem seča prizadeti tudi

<sup>1</sup>[www.dlvkos.si](http://www.dlvkos.si)

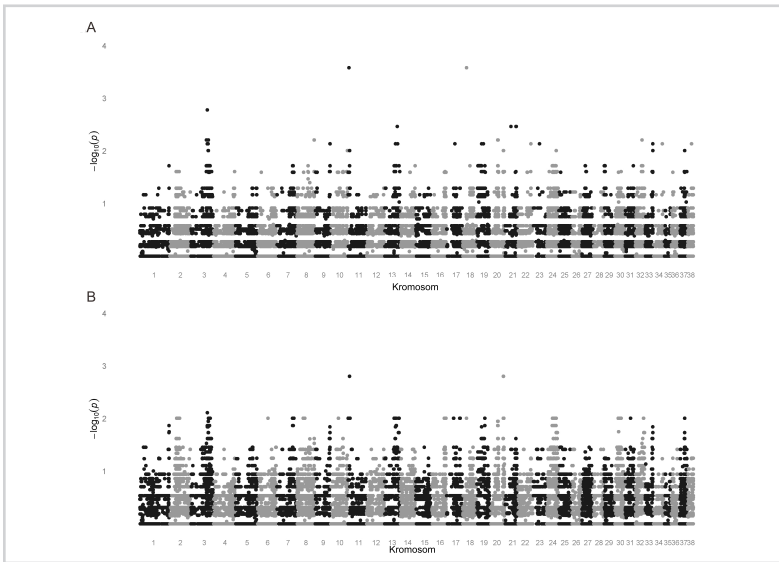
<sup>2</sup>[www.genopro.com](http://www.genopro.com)

psi križanci kraških ovčarjev, kar namiguje na nedefinirano raven dominantnega dedovanja. Najbližji skupen prednik (angl. the most recent common ancestor, TMRCA) psov z uhajanjem seča je enak kot za celotno populacijo. To je psica, skotena leta 1968, kar govori proti enostavnemu enogenskemu dedovanju. Dokumentiran je tudi primer prizadete samice s potomci brez znakov boleznih sečil, pri čemer pa se je ektopija sečevodov pojavila v generaciji F<sub>2</sub>. Poleg tega analiza rodovnikov kaže, da nekateri psi z boleznimi sečil oz. njihovi bližnji sorodniki izkazujejo znake nekaterih drugih boleznih, kar kaže na kompleksno gensko ozadje in vlogo genov, ki spreminjajo izražanje drugih genov (angl. modifier genes). Tudi v primeru, če bi bila za ektopijo sečevodov odgovorna mutacija enega gena, je precej verjetno, da na penetranco in raznoliko izrazitost obolenja vplivajo še drugi dejavniki.

### 3.2.3 *Asociacijska preiskava celotnega genoma in kartiranje homozigotnosti*

DNA smo izolirali iz vzorcev krvi petih prizadetih (en samec, štiri samice) in petih psov brez kliničnih znakov (trije samci, dve samici). Vzorci so vključevali trojico: samca in samico brez kliničnih znakov ter njunega prizadetega potomca. Vsi prizadeti psi so imeli potrjeno diagnozo veterinarja, vsi psi brez kliničnih znakov pa so bili starejši od štirih let in niso imeli zgodovine boleznih sečil. Vzorci DNA so bili hibridizirani na SNP-čip CanineHD Beadchip (Illumina, 9885 Towne Centre Drive, San Diego, CA 92121, ZDA) v genomskem centru kalifornijske univerze Davis (DNA Technology Core, UC Davis Genome Center). Asociacijsko preiskavo celotnega genoma smo opravili s programskim paketom Plink v 1.07 [52] in pri tem uporabili alelni test ter Cochran-Armitage test. Po izločanju SNP-jev z več kot 10% stopnjo manjkajočih genotipov ter SNP-jev s pogostostjo manj pogostega alela v populaciji (angl. minor allele frequency, MAF), ki je nižja od 0,01, je v analizi ostalo 98 723 informativnih SNP-jev. Celokupna stopnja genotipizacije za vse osebke je bila > 0,99. Genomski dejavnik inflacije (glede na vrednost mediane hi-kvadrat) z vrednostjo 1 potrjuje odsotnost populacijske razslojenosti v analiziranih vzorcih. Z uporabo alelnega testa smo odkrili povezavi na kromosomih CFA11 (rs8971641) in CFA18 (BICF2S3655583) ( $P_{raw} = 0,00026$ ) (slika 3.4 A). SNP rs8971641 je v območju introna gena COL23A1 (CanFam3.1, CFA11: 3 242 754), SNP BICF2S3655583 pa je 23 kb pred začetkom gena MLL5 (CanFam3.1, CFA18: 15 331 404). Z uporabo Cochran-Armitage testa smo najmočnejše povezave odkrili na kromosomih CFA11 in CFA20 ( $P_{raw} = 0,0016$ ), za katere so bili vsi prizadeti psi v homozigotnem in vsi psi brez znakov boleznih v hetero-

zigotnem stanju (slika 3.4 B in slika 3.8 B).

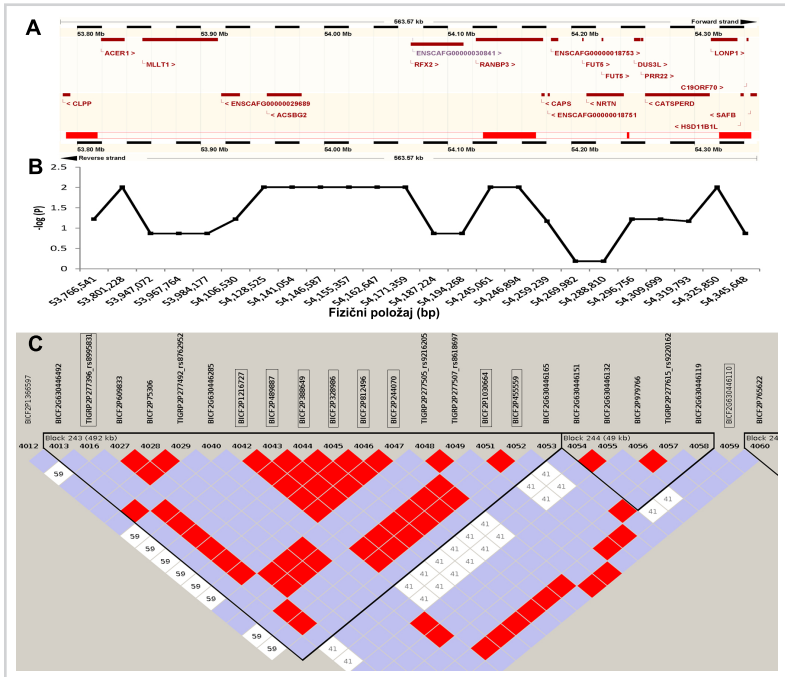


Slika 3.4

Grafa asociacijske preiskave celotnega genoma za ektopijo sečevodov, ki predstavlja  $-\log_{10}$  surovih P-vrednosti za vsak SNP v analizi obolelih proti psom brez kliničnih znakov z (A) alelnim in (B) Cochran-Armitage testom.

Vepravno neravnovesje smo preiskovali s štiri-gametnim pravilom v programu Haploview [53]. Dva sosednja SNP-ja na CFA11 sta opredelila 214 kb območje povezave (CanFam3.1, CFA11: 5 838 742, CFA11: 6 011 599), ki prekriva gen COMMD10. Deset SNP-jev na CFA20 je opredelilo štiri sklope (CanFam3.1, CFA20: 53 791 049–53 816 416, CFA20: 54 128 525–54 171 359, CFA20: 54 245 061–54 246 894, CFA20: 54 319 793–54 345 648), ki so znotraj dveh blokov vezavnega neravnovesja (slika 3.5). Prvi sklop prekriva gen *CLPP*, drugi prekriva gen *RANBP3*, tretji je 4 kb stran od začetka gena *NRTN* in četrti prekriva gene *LONP1*, *HSD11B1L* in *ci9orf70*. SNP-ji z najbolj značilnimi signali in njihovi genotipi v obolelih psih in psih brez kliničnih znakov so prikazani spodaj (slika 3.6).

Gene iz odkritih območij in njihove povezave s fenotipom sečil navajamo v spodnji tabeli (tabela 3.1). Zaradi problematike testiranja povezav posameznih lokusov smo dodatno uporabili večlokusni mešani model (MLMM) [54] v programu SNP & Variation Suite (SVS) 7 (Golden Helix Inc., Bozeman, MA, ZDA) in z visoko statistično značilnostjo potrdil vse ugotovljene povezave, razen SNP-ja na CFA18.



Slika 3.5

(A) Območje odkrite povezave z anotacijami genov na kromosomu CFA2o. (B) Deset SNP-jev na CFA2o, ki določajo štiri signale povezane in so znotraj (C) dveh blokov vezavnega neravnovesja.

Slika 3.6

Štirinajst značilnih SNP-jev in njihovi genotipi pri psih brez znakov bolezni in pri prizadetih psih.

CFA	SNP	PSI BREZ ZNAKOV BOLEZNI								PSI Z EKTOPIČNIM SEČEVODOM									
11	TIGRP2P1500650_r88971641	A	A	A	A	G	C	A	A	A	A	A	C	C	C	C	C	C	C
11	BICF2S2293721	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T
11	BICF2P1125455	G	T	G	T	G	T	G	T	T	T	T	T	T	T	T	T	T	T
18	BICF2S23655583	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C
20	TIGRP2P277396_r88995831	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T
20	BICF2P1216727	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T
20	BICF2P489867	G	T	G	T	G	T	G	T	G	T	G	T	G	T	G	T	G	T
20	BICF2P388649	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G
20	BICF2P328986	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G
20	BICF2P812496	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T
20	BICF2P244070	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T
20	BICF2P1030664	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T
20	BICF2P455559	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T
20	BICF2G830446110	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C

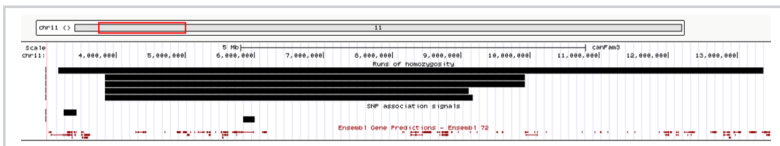
Tabela 3.1: Kandidatni geni za nagnjenost k ektopiji sečevodov.

Simbol gena	Območje (bp) v CanFam3.1	Povezava s fenotipom sečil
<i>COL23A1</i>	11: 3048578–3320226	visoko izražanje gena v medceličnem prostoru ledvične sredice [55]
<i>MLL5</i>	18: 15240543–15308087	visoko izražanje gena v mezenhimu razvijajočih ledvic okrog konic sečevodovega brstiča [55]
<i>COMMD10</i>	11: 5800336–5984594	nepoznana
<i>CLPP</i>	20: 53788312–53794384	nepoznana
<i>RANBP3</i>	20: 54122903–54177083	znižano izražanje gena v zgodnjem razvoju ledvic med E10,5 in E13,5 v mišjem modelu [56]
<i>NRTN</i>	20: 54212282–54242456	avtokrini rastni faktor, ki uravnava razvoj ledvičnih zbirnih vodov [57]
<i>LONP1</i>	20: 54313027–54334405	nepoznana
<i>HSD11B1L</i>	20: 54337038–54339544	visoko izražanje gena v ledvicah [58]
<i>ci9orf70</i>	20: 54342064–54343467	nepoznana

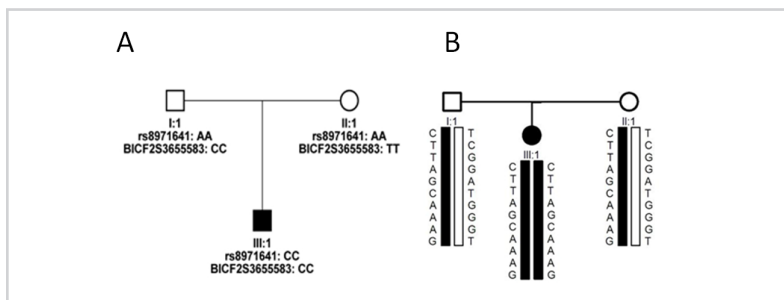
Kartiranje homozigotnosti smo izvedli s Plink ukazom “-homozyg-density 100 – homozyg-group” in tako odkrili 5,27 Mb kritično območje na kromosomu CFA11, ki se razprostira preko signalov povezav in je homozigotno pri vseh prizadetih psih in nobenem od psov brez znakov bolezni (slika 3.7).

Slika 3.7

Kartiranje območja homozigotnosti na kromosomu CFA11 pri petih psih z ektopijo sečevodov. Vsi prizadeti psi (in nobeden od psov brez kliničnih znakov) so homozigotni za 5,27 Mb kritični odsek, ki se razteza preko signalov povezav.



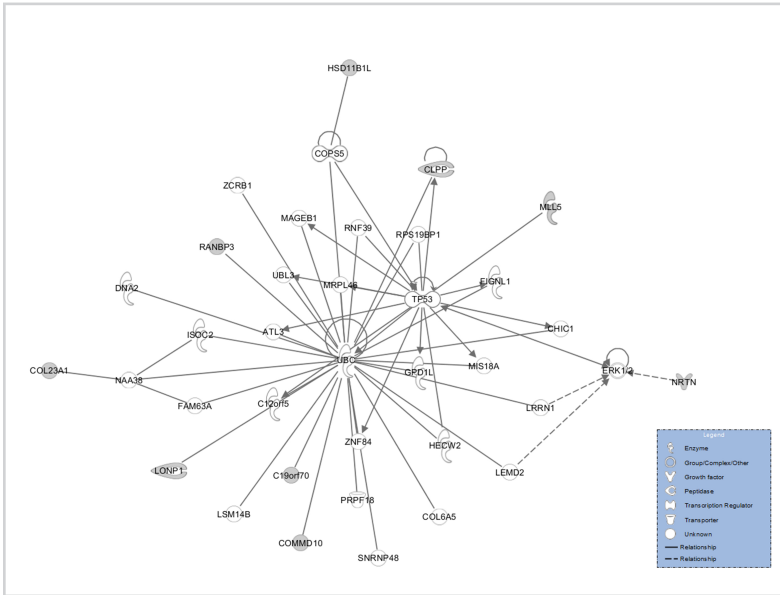
S programom Plink smo preverjali tudi mendelska neskladja genotipov. Preverjanje za mendelska neujemanja je razkrilo 167 napak. Od tega dve mendelski neujemanji ustrežata signaloma SNP-jev rs8971641 na kromosomu CFA11 in BICF2S3655583 na kromosomu CFA18 (slika 3.8 A). Z matriko podobnosti IBS (angl. identity-by-state) smo potrdili, da je oboleli pes trojice najbolj podoben svojim staršem in zato mendelsko neskladje najverjetneje ni posledica nenamerne zamenjave vzorcev.



Slika 3.8

(A) Mendelska neskladja dveh SNP-jev, odkritih z alelnim testom. (B) Deset SNP-jev na CFA20, heterozigotnih pri starših in homozigotnih pri potomcu in odkritih z Cochran-Armitage testom.

Kandidatne gene iz območij signalov smo uporabili kot vhodne podatke v analizi funkcij genov z uporabo programa Ingenuity Pathway Analysis Software (Ingenuity Systems, Inc., Redwood City, CA, ZDA). Kandidatne gene smo analizirali glede na funkcije njihovih ortologov pri človeku. Najbolj značilni biološki procesi so bili celični cikel, embrionalni razvoj ter podvojevanje DNA, rekombinacija in popravilo, vsi kandidatni geni pa so bili združeni znotraj skupne genske mreže (slika 3.9).

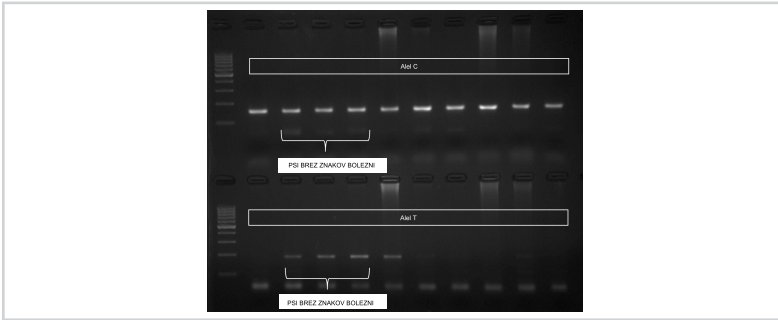


Slika 3.9

Mreža kandidatnih genov (prikazani s sivo barvo) iz območij signalov povezav. Mreža je obogatena za biološke procese celičnega cikla, embrionalnega razvoja in podvojevanje DNA, rekombinacije in popravila.

Za 14 najbolj značilnih zadetkov (slika 3.6) smo preverili pogostost genotipov v dodatnem kontrolnem naboru genotipov 409 psov pasem veliki pudelj in bradati škotski ovčar. Odkrili smo raznolike porazdelitve genotipov s prisotnimi homozigotnimi ter heterozigotnimi genotipi za 6 SNP-jev. Presenetljivo je, da pri ostalih 7 SNP-jih homozigotnih genotipov za alele prizadetih kraških ovčarjev nismo odkrili v nobenem od velikih pudljev ( $n=122$ ) ali bradatih škotskih ovčarjev ( $n=287$ ). Šest od teh SNP-jev se je znotraj enega bloka vezavnega neravnovesja. Preučitev 14 najbolj značilnih SNP-jev v populaciji kraškega ovčarja je predmet nadaljnjih raziskav. Začetni rezultati namigujejo, da nobeden od SNP-jev na kromosomih CFA11, CFA18, niti SNP BICF2P812496 na CFA20 ne razlikuje primerov brez znakov bolezni od primerov z ektopijo sečevodov v popolnosti. Kot primer prikazujemo spodaj (slika 3.10) gel agarozne elektroforeze določanja alelov SNP-ja BICF2P812496 v dodatnem vzorcu kraških ovčarjev z diagnozo ektopije sečevodov z metodo ARMS (angl. Amplification Refractory Mutation System). Reakcije PCR smo izvedli v namiznem mikroprocesorsko vodenem termostatu GeneAmp PCR System 2720 (AB Applied Biosystems). Zaporedji zunanjih začetnih oligonukleotidov sta bili 5'-CTGCTTGTAGAGGAAAAGAGCAGTGGAT-

3' in 5'-CCTCAGGTAGGACAATCTTCTTGGTCTG-3', zaporedji notranjih začetnih oligonukleotidov pa 5'-AAGGAAAGTGGAAGGGCTTTCTCGAT-3' za alel T in 5'-GTGCAAACAGATGCTACACAGAGAGTTG-3' za alel C. Genomsko DNA (približno 50 ng) smo pomnožili v končni prostornini 10 µl, ki je vsebovala 10 pmolov vsakega začetnega oligonukleotida, 100 µM mešanico dNTP-jev, 2,5 mM MgCl<sub>2</sub>, 1 mM PCR pufra in 0,5 enote Taq polimeraze (Fermentas). Po dvominutni začetni denaturaciji pri 95°C je sledilo 35 ciklov pomnoževanja, ki so obsegali 60 sekund denaturacije pri 95°C, 60 sekund prileganja začetnih oligonukleotidov pri 68°C in 60 sekund podaljševanja pri 72°C, ki jim je sledilo še desetminutno podaljševanje pri 72°C. Produkta PCR smo analizirali na 2,5% agaroznem gelu. Genotipizacija tega SNP-ja je pokazala, da vloge tega območja pri razvoju ektopičnega sečevoda ne moremo trajno potrditi. Poskusi genotipizacije večjega števila SNP-jev s kompetitivno alelno-specifično reakcijo PCR (KASP by Design, LGC Genomics) so bili tehnično neuspešni, količina DNA vzorcev psov z diagnozo pa omejena za večje število poskusov izboljšanja učinkovitosti reakcije. Eksonom kandidatnega gena *NRTN* smo poskusili določiti zaporedje kraševcev brez bolezenskih znakov ter kraševcev z boleznijo, vendar je bilo kljub preizkusom različnih pogojev PCR sekvenciranje neuspešno oz. razmerje signala proti šumu prenizko za zanesljiv rezultat. V času naših preiskav je bila ta gen v genomu CanFam2.0 anotiran kot podvojitev, med paralogoma tega gena pa je bila prisotna vrzel v referenčnem zaporedju. Poleg tega so se razlikovale tudi anotacije eksonov v različnih podatkovnih zbirkah (Ensembl in NCBI). V trenutni anotaciji pasjega genoma (CanFam3.1) je anotirana ena kopija gena *NRTN*, spremenile so se tudi anotacije eksonov, a so v zaporedju referenčnega gena še vedno prisotne vrzeli. Za dokončno določitev vloge kandidatnih genov bo pomembna tudi nadaljnja izboljšava anotacije genoma psa.



Slika 3.10

Genotipizacija SNP-ja BICF<sub>2</sub>P812496 v dodatnih primerih ektopije sečevodov. Večina obolelih psov ima genotip CC, vendar pa je bil v analizi prisoten tudi primer pسية z ektopičnim sečevodom z genotipom CT, kar kaže na to, da odkriti SNP ne razlikuje primerov brez bolezenskih znakov popolnoma od primerov z boleznijo.

### 3.3 Diskusija in prihodnji izzivi

V tem poglavju predstavljamo začetne rezultate dolgoročneje raziskave ektopije sečevodov in sindroma uhajanja seča pri pasmi kraški ovčar. Rodovniški rezultati so skladni s kompleksnim genskim ozadjem bolezni kot tudi epidemiološke in klinične značilnosti bolezni pri tej pasmi. Analiza rodovnikov nakazuje, da je najbolj verjetna razlaga za raznolike fenotipske izrazitosti bolezni sečil pri kraševcih večgensko dedovanje, oziroma je v primeru enogenskega dedovanja pomembna vloga genov, ki vplivajo na izražanje drugih genov. Kot je razvidno iz rodovnika ene od družin kraških ovčarjev, se lahko znotraj posameznih legel pojavljajo bolezni sečil, mrtvorojeni mladiči pa tudi psi brez znakov bolezni. Med obolelimi živalmi prevladujejo samice, samci z ektopijo sečevodov pa so odkriti bolj redko in običajno kasneje kot samice. Najverjetnejša razlaga je, da je moška sečnica daljša in dovoljuje povratno polnjenje sečnega mehurja [59]. Poleg tega se dednost ektopije sečevodov prepleta z genskim ozadjem več drugih dednih bolezni pri tej pasmi. Z namenom, da bi odkrili kandidatne lokuse, ki vplivajo na nagnjenost pasme k temu obolenju, smo opravili asociacijsko preiskavo celotnega genoma. Z genotipizacijo petih prizadetih ter petih psov brez znakov bolezni na 170k SNP-čipu smo odkrili več položajskih in funkcijskih kandidatnih genov na kromosomih CFA11, CFA18 in CFA20 ter predlagali kompleksen vzorec dedovanja, ki smo ga predpostavili že na osnovi analize rodovnikov. Z iskanjem mendelskih neskladij, kartiranjem območij homozigotnosti ter uporabo genotipskih podatkov drugih pasem smo dodatno potrdili možne lokuse, ki vplivajo na nagnjenost k obolenju. Ta raziskava predstavlja začetna odkritja, da je v pojav ektopičnega sečevoda pri kraškem ovčarju najverjetneje vpleteno več genov iz skupnih bioloških poti celičnega cikla in

embrionalnega razvoja. Pomanjkljivost raziskave je, da so bili kontrolni psi izbrani naključno, na podlagi odsotnosti kliničnih znakov. Veterinarski pregled sečil za te pse pred vključitvijo v analizo ni bil opravljen. Da sta dva od petih kontrolnih psov dala potomca z ektopičnim sečevodom v generaciji F<sub>1</sub>, eden pa potomca z ektopičnim sečevodom v generaciji F<sub>2</sub>, smo odkrili kasneje in nato testirali mendelsko skladnost. V praksi je diagnostika ektopije sečevodov problematična in pogosto ni nedvoumno potrjena tudi v primerih uhajanja seča pri tej pasmi. Čeprav nobeden od ugotovljenih SNP-jev ni bil statistično značilen po prilagoditvi za večkratno testiranje, je za to lahko odgovoren tudi majhen vzorec. Tako imenovan problem "velik p, majhen n" je že poznana težava asociacijskih preiskav celotnega genoma. Ker za obolenje ni nujno odgovoren en sam lokus, ampak je zaradi populacijske strukture in vezavnega neravnovesja lahko z danim stanjem povezanih več vzročnih alelov, smo dodatno uporabili še večlokusni mešani model (MLMM) in potrdili skoraj vse odkrite povezave z visoko statistično značilnostjo. Testiranje mendelskih neskladij je razkrilo, da sta genotipa dveh najmočnejših signalov, ugotovljenih z alelnim testom, pri potomcu nezdružljiva z genotipi staršev. Matrika IBS je potrdila, da napake niso posledica naključne zamenjave vzorcev in nakazala, da so poleg naključnih napak genotipizacije možen vzrok tudi različice v številu kopij (angl. copy-number variation, CNV) - npr. odkritje hemizigotne delecije kot homozigotnega stanja. Oba kandidatna gena, *COL23A1* in *MLL5*, sta značilno diferencialno izražena med razvojem ledvic in sečevodov [55]. Raziskave SISH (angl. high resolution in situ hybridization) razvoja ledvice v mišjem modelu od embrionalnega dneva (E) 15,5 naprej so dokazale obogateno izražanje *COL23A1* v medceličnini ledvične sredice in obogateno izražanje *MLL5* v ledvičnem mezenhimu okrog konic sečevodovega brstiča. Poleg tega dodatno podpirajo signal na kromosomu CFA11 rezultati kartiranja območij homozigotnosti. Odsotnost genotipov prizadetih psov v 409 kontrolnih psih podpira odkritje sedmih SNP-jev na kromosomu CFA20 in ponuja razlago z usmerjeno selekcijo proti tem genotipom ali njihovo gensko povezanost z drugimi genotipi, ki vplivajo na fitnes živali. Medtem ko gen *NRTN* kodira avtokrini rastni faktor, ki uravnava razvoj ledvičnih zbirnih vodov [57], drugi geni še nimajo znane specifične funkcije pri razvoju sečil. Kljub temu vsi odkriti geni medsebojno vplivajo in uravnavajo celični cikel in poti razvoja organov in so tako možni kandidatni geni za ektopijo sečevodov. Čeprav je bilo število vzorcev v tej analizi premajhno za prepričljive dokaze o načinu dedovanja, rezultati namigujejo, da ektopija sečevodov ni preprosta enogenska bolezen, temveč k penetranci in nagnjenosti za to

obolenje pri kraškem ovčarju prispeva več genov. Zato je verjetno, da bo za analizo genske nagnjenosti k razvojnim obolenjem sečil pri tej pasmi v prihodnje potrebna raziskava v večjem merilu. Za primerjavo naj omenimo, da so asociacijsko analizo celotnega genoma te bolezni že opravili v okviru raziskav ektopije sečevodov pri pasmi entlebuški planšarski pes in z vključitvijo 48 psov z ektopijo sečevodov in 48 kontrolnih psov odkrili povezavo na kromosomu CFA7 ( $P = 0,088$ ), vendar pa nameravajo raziskovalci za večjo zanesljivost rezultatov analizo opraviti na okvirno 300 živalih (I. Reichler, osebna komunikacija). Pri kraševcu bi bila koristna natančnejša epidemiološka in fenotipska opredelitev različnih oblik te bolezni. Glede na nizko penetranco bolezni bi bili smiselni presejalni testi populacije kraševca po zgledu raziskav pri entlebuškem planšarskem psu. Pri tej pasmi razlikujejo in razvrščajo fenotip kot normalna ustja sečevodov v trikotniku mehurja (EU-A), ektopična ustja v področju vratu sečnega mehurja (EU-B) in ekstravezikularne ektopične sečevode (EU-C). Veliko pogostost bolezni pri tej pasmi so ugotovili predvsem na račun primerov EU-B, ki pa so v smislu razlikovanja od fiziološko normalnih primerov EU-A diagnostično težavni. Še posebej zato, ker diagnostična opredelitev temelji na meritvah razdalj ustij sečevodov in sečnice, ki pa so močno odvisne od polnosti sečnega mehurja, velikosti psa in referenčne mejne vrednosti, ki jo predpostavimo kot fiziološko še normalno. Zato so v tej raziskavi uporabili kontrolno skupino apencelskih planšarskih psov, pasme, pri katerih je ektopični sečevod manj pogost, fizionomija pa podobna tisti pri entlebuških planšarskih psih. Na podlagi ugotovljenih položajev ustij sečevodov so bila oblikovana vzrejna pravila. Nemško društvo entlebuških planšarskih psov je na primer iz vzreje izločilo pse z diagnozo EU-C, pse z diagnozo EU-B pa se lahko pari v kombinaciji s psi z diagnozo EU-A. Paritve dveh psov z diagnozo EU-B so dovoljene le ob zavezi vzreditelja, da poskrbi za veterinarski pregled sečil celotnega legla v starosti šestih do dvanajstih mesecev ter društvu preda podatke o novih lastnikih mladičev, saj klinični pomen teh paritev še ni znan. Izvedba takšnega vzrejnega programa je od leta 2008 že občutno znižala pojavnost ektopičnega sečevoda pri tej pasmi (R. Fritsche, osebna komunikacija). Z nadaljnjim zbiranjem tkivnih vzorcev, ki smo ga začeli zbirati ob primerih evtanaziranih kraševcev z ektopijo sečevodov, bi v prihodnosti bile možne tudi histopatološke raziskave ter raziskave izražanja genov ali epigenetike. V začetni raziskavi smo tako pripravili osnovo za nadaljnje genske raziskave ektopičnih sečevodov pri kraškem ovčarju. Preučitev odkritih SNP-jev in uporaba dodatnih analitičnih postopkov (npr. komparativna genomska hibridizacija na čipih, sekvenciranje celotnega genoma

ali tarčnega območja) ponujajo možnosti za nov vpogled v razvojna obolenja sečil pri psih. Odkritje vzročnih mutacij bi omogočilo boljše razumevanje fiziologije te bolezni, določitev pogostosti alelov v širši populaciji kraševca, primerjavo vzročnih mehanizmov pri drugih pasmah, nagnjenih k ektopiji sečevodov, potrjevanje vzročnega mehanizma z dodatnimi preiskavami (npr. celičnim modelom ali analizo na ravni beljakovin) ter ne nazadnje učinkovitejšo diagnostiko, zdravljenje in preprečevanje tovrstnih obolenj z ustreznimi rejskimi ukrepi.

*Analiza transkriptoma celic  
mlečne žleze kože po okužbi z  
bakterijo *Mycoplasma  
agalactiae* in vitro*

#### 4.1 Ozadje

Do nedavnega so bile raziskave izražanja večjih naborov genov omejene le na modelne organizme, vendar pa je hiter razvoj visokozmogljivostnega sekvenciranja sprožil preobrat na področju ugotavljanja transkriptomov. Omogočil je natančno in cenovno ugodno določanje, ugotavljanje značilnosti in količine izraženih prepisov ter določanje vzorcev spajanja in tako uveljavil sekvenciranje RNA kot zelo ustrezno tehniko za preučevanje transkriptomov praktično vseh bioloških vrst.

V javnih podatkovnih zbirkah je še vedno omejena količina podatkov o izražanju genov koze (*Capra hircus*). Vendar pa se to dejstvo hitro spreminja, saj se tudi koza hitro uveljavlja kot pomemben model na področjih biomedicine in molekularne biologije [60–62]. Pred kratkim je bilo objavljeno zaporedje referenčnega genoma koze [63], ki odpira še dodatne možnosti za raziskovanje genoma koze na različnih raziskovalnih področjih.

Kužna agalakcija je ena od gospodarsko najpomembnejših bolezni drobnice v svetovnem merilu. Pri kozah je glavni povzročitelj, ki je odgovoren za približno 90% vseh izbruhov bolezni, bakterija *Mycoplasma agalactiae* [64], vendar pa je znanje o zgodnjih stopnjah gostiteljevega odziva še vedno skromno. Klinično se bolezen po okužbi mlečne žleze lahko izrazi kot nenadni mastitis, vendar pa običajno napreduje v zmerno obliko. Klinični znaki vključujejo upad ali popolno prenehanje izločanja mleka, možen je tudi razvoj artritisa in keratokonjunktivitisa. Mikoplazma se sprošča v mleko še dolgo po tem, ko se klinični znaki izboljšajo.

*M. agalactiae* spada v razred *Mollicutes*, za katerega je značilna odsotnost celične stene, skromen genom in omejene metabolne sposobnosti, saj se za svoje preživetje zanaša na tesen stik z gostiteljevimi epitelnimi celicami. O njenih virulentnih dejavnikih in medsebojnem vplivanju med mikoplazmo in gostiteljem je še vedno malo znanega. Tovrstno znanje bo ključno za nadzor, zdravljenje in preprečevanje bolezni, pri kateri je uporaba klasičnih antibiotikov neučinkovita, uporaba cepiv, ki so trenutno na voljo, pa zaenkrat še vedno tudi le delno uspešna. Ker lahko določitev nabora prepisov, ki se izražajo med okužbo celic mlečne žleze koze, olajša razumevanje fiziologije in mnogoličnosti odnosov med gostiteljem in bakterijo, posredno pa pripomore k učinkovitejšim selekcijskim in zdravstvenim ukrepom, smo se v tej raziskavi z analizo podatkov prepisovanja RNA okuženih celic *in vitro* posvetili izboljšanju anotacije genoma koze in vpogledu v odziv celic mlečne žleze koze na okužbo z *M. agalactiae*.

## 4.2 Metode in rezultati

### 4.2.1 Opis eksperimentalnih podatkov

Aanaliza transkriptoma temelji na laboratorijskem delu Ogorevca in sodelavcev [65, 66]. Njihovi surovi podatki so dostopni v podatkovni zbirki Gene Expression Omnibus z oznako GSE30379. V tej raziskavi so po prilagojenem protokolu [67] vzpostavili celično kulturo, obogateno z epitelno celično populacijo iz približno 500 g začetnega materiala mlečne žleze sanske koze v laktaciji. Celice so nasadili v rastni medij RPMI-1640 z dodanimi amino kislinami (metionin (0,1 mM), lizin (0,4 mM), glutamin (2mM)), natrijevim bikarbonatom (2g/L), natrijevim piruvatam (0,01mM), 10% zarodnim govejim serumom ter mešanico hormonov, ki spodbujajo nastajanje in izločanje mleka (inzulin (1 µg/ml), prolaktin (1 µg/ml), hidrokortizon (1 µg/ml)) na plastično podlago prevlečeno z beljakovinskim izvlečkom bazalne membrane (Geltrex, Life Technologies). Po prvi pasaži so bile celice gojene v celični kulturi do 90% preraščenosti dna gojitvene posodice (konfluence) (ca.  $8 \times 10^6$  celic/  $75 \text{ cm}^2$ ). Bakterijsko kulturo *Mycoplasma agalactiae* so pripravili kot opisano v [68]. Po 13 dneh od vzpostavitve celične kulture mlečne žleze so bile celice okužene z *M. agalactiae* (sev PG2), nato pa so s komercialnim kitom (RNA miRNeasy kit, Qiagen) izolirali celokupno RNA v štirih časovnih točkah: 0h, 3h, 12h in 24h po okužbi. Vse analize so opravili v trojnikih. Količino RNA so določili spektrofotometrično (nanoVue, GE Healthcare Life Sciences), v postopku priprave knjižnice za sekvenciranje pa so uporabili komercialna kita mRNA-Seq Sample Prep Kit (Illumina) in Multiplexing Sample Preparation Oligonucleotide Kit (Illumina). Po sintezi komplementarne DNA (cDNA) in ligaciji oligonukleotidnih adapterjev so iz gela izolirali fragmente dolžine 200 baznih parov (bp). Koncentracija knjižnice cDNA so določili s fluorometrom Qubit (Invitrogen), velikost in čistost pa z aparatom Agilent 2100 Bioanalyzer in komercialnim kitom Agilent DNA 1000 Kit. Vseh 12 vzorcev (4 časovne točke po 3 vzporedne ponovitve) so hkrati sekvencirali petkrat. Sekvenciranje RNA so izvedli na sekvencatorju druge generacije Illumina Gene Analyzer IIx. Podatki so zapisani v formatu fastq (phred+64). Kakovost določenih zaporedij vseh petih prog smo preverili in potrdili s programskim paketom ShortRead [69]. V celoti je bilo določenih 67 589 924 zaporedij dolžine 50 bp.

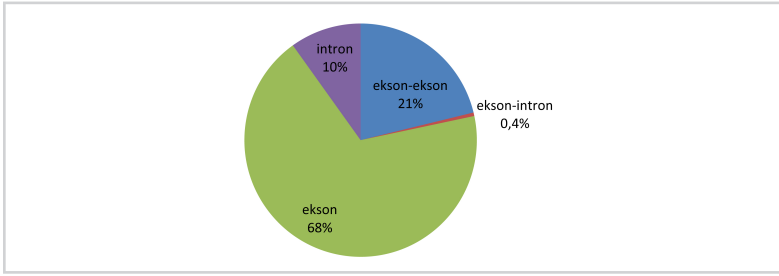
## 4.2.2 Kartiranje zaporedij na referenčni genom

Kratke odčitke RNA smo kartirali na referenčne genome kože (*Capra hircus*), goveda (*Bos taurus*), ovce (*Ovis aries*) in bakterije *M. agalactiae* seva PG2.

Z uporabo algoritma programskega paketa CLC Genomics Workbench (CLC Bio, Aarhus, Danska) smo kartirali zaporedja na genome goveda (Btau4.0), ovce (OviAri1) in *M. agalactiae* (ASM6360v1). Pri tem smo uporabili privzete parametre za kazen neujemanja (angl. mismatch), ki je bila enaka 2, in kazen insercij in delecij, ki je bila enaka 3. Najmanjši odsek odčitka, ki se je moral ujemati z referenčnim zaporedjem (angl. length fraction), je bil enak 0,5. Meja identitete tega odseka z referenčnim zaporedjem (angl. similarity) je bila postavljena na 80%. Od vseh 68 milijonov določenih zaporedij se je 48 745 998 (72,1%) zaporedij prilegalo na genom ovce in 41 942 390 (62,1%) na genom goveda (tabela 4.1), kar kaže na večjo podobnost ortolognih zaporedij ovce in kože in je skladno z bližjo sorodnostjo ovce in kože glede na molekularno filogenetske analize [70–73]. Kljub ekstrakciji informacijske RNA (mRNA) smo zaznali izražanje 473 genov *M. agalactiae* (59,6% od 794 anotiranih). Odčitkom, ki smo jih kartirali na referenčni genom goveda, smo določili gensko lokacijo. Večina zaporedij je kartirala v celoti v eksonih (68%), hkrati pa smo 21% zaporedij odkrili na mejah med eksoni, 10% v intronih in 0,4% na mejah med introni in eksoni (slika 4.1). Odčitki, ki kartirajo v intronih, lahko predstavljajo eksone, ki v Btau4.0 niso anotirani, ali pa primarne prepise. Odčitki, ki kartirajo na mejah eksonov, so znak alternativnega spajanja eksonov, odčitki, ki kartirajo na mejah med introni in eksoni, pa najverjetneje predstavljajo primarne prepise.

Tabela 4.1: Število in delež odčitkov, ki je kartiral na različne genome.

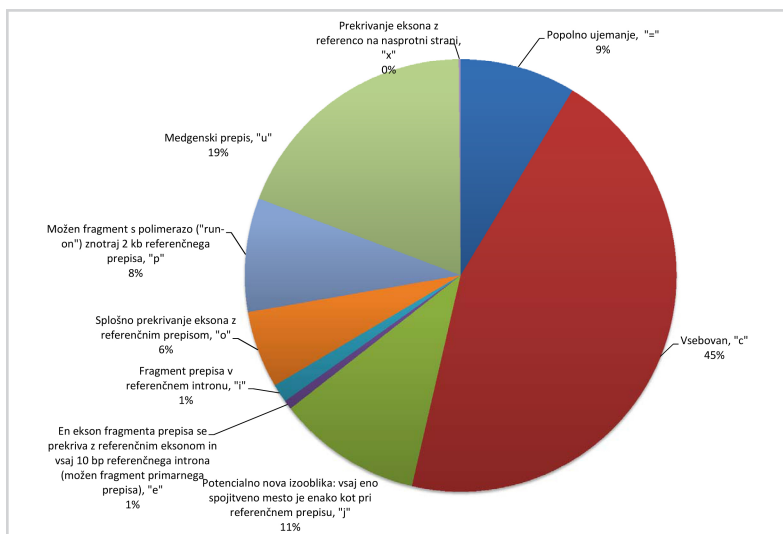
Vrsta	<i>Bos taurus</i>	<i>Bos taurus</i>	<i>Ovis aries</i>	<i>Capra hircus</i>
Algoritem	CLC Bio	TopHat2	CLC Bio	TopHat2
Združek genoma	Btau4.0 (NCBI)	UMD3.1 (Ensembl)	OviAri1.0 (NCBI)	CHIR1.0 (NCBI)
Št. genov v anotaciji	27 368	26 719	/	21 389
Št. kartiranih odčitkov	41 942 390	42 768 101	48 745 998	55 585 971
% vseh odčitkov	62,1%	63,3%	72,1%	82,3%



Slika 4.1

Porazdelitev kartiranih odčitkov na različna področja genomov.

Z uporabo odprtokodnih orodij zbirke Tuxedo smo odčitke kartirali tudi na genoma koze (CHIR1.0) in goveda (UMD3.1). Z orodjem TopHat2 [74] smo zaporedja najprej poravnali na referenčne genome z algoritmoma Bowtie 1 [75] in Bowtie 2 [76], ki uporabljata Burrows-Wheelerjev indeks genoma za hitro poravnavo kratkih zaporedij. TopHat na podlagi poravnave ocenjuje najverjetnejše meje med eksoni in introni in odkrije zaporedja, ki prilegajo na meje. Kazen za neujemanja je znašala 2, kazni za insercije ter deležije pa 3. Tudi ostalim parametrom kartiranja odčitkov nismo spreminjali privzetih vrednosti. Na genomu koze je z algoritmom Bowtie 1 kartiralo 55 157 565 (81,7%) zaporedij, z algoritmom Bowtie 2 pa 55 585 971 (82,3%) zaporedij. Na genomu goveda (UMD3.1) je z algoritmom Bowtie 1 uspešno kartiralo 41 645 385 (61,7%) zaporedij, z algoritmom Bowtie 2 pa 42 768 101 (63,3%) zaporedij. Algoritem Bowtie 1 je bil izdan leta 2009 in je namenjen predvsem poravnavi relativno kratkih odčitkov (do 50 bp) in iskanju poravnave brez vrzeli. Nasprotno pa Bowtie 2 podpira poravnavo zaporedij z vrzeli in temu prilagaja kazni vrzeli. Ker so bili v tej raziskavi odčitki dolgi 50 bp, smo uporabili oba algoritma. Po pričakovanjih sta oba algoritma dala primerljive rezultate (tudi v primerjavi s CLC Bio), vendar smo v nadaljnjih korakih analize uporabili poravnave z Bowtie 2, saj je v primeru obeh genomov kartiral nekoliko več odčitkov. Za spajanje anotacij združkov (angl. assembly) transkriptoma, primerjavo zaznanega prepisovanja z obstoječimi anotacijami genomov in ugotavljanje razlik izražanja genov med različnimi časovnimi točkami smo uporabili dodatne programe iz paketa Cufflinks [77]. Združevanje sosesk (angl. contigs) v nove in znane izooblike smo podprli z datotekami anotacij referenčnih genomov in s tem izboljšali kakovost združevanja zaporedij. Na ta način smo pripravili anotacijo celotnega transkriptoma, ki je predstavljala 58 593 fragmentov prepisov. Od teh se 9 % popolnoma ujema z anotacijo genoma *C. hircus* (slika 4.2), 19 % prepisov je med geni, 11 % pa jih predstavlja



Slika 4.2

Porazdelitev različnih tipov prepisov in fragmentov prepisov.

Z namenom iskanja možnosti izboljšave trenutne strukturne anotacije genoma koze smo s programskim paketom Cufflinks pridobili skupino fragmentov prepisov, zgrajenih iz odčitkov, ki glede na poravnavo na genom predstavljajo soseske, vendar so med geni (označeni so z razredno kodo "u"), v intronih (označeni so z razredno kodo "i"), ali predstavljajo nove anotacije eksonov (označeni so z razredno kodo "o") oziroma nove izooblike (označeni so z razredno kodo "j"). Iz izhodne datoteke anotacije smo razčlenili vse fragmente prepisov, označene z razrednimi kodami "u", "i", "o" in "j". Zbranih je bilo 8 732 novih prepisov, ki smo jih nato uporabili za napovedovanje odprtih čitalnih okvirjev z orodjem TransDecoder<sup>1</sup>. Za 6 110 (70%) zaporedij smo določili, da predstavljajo dobra kandidatna kodirajoča območja. Z namenom anotacije zaporedij smo z algoritmom blastn poiskali njihove ortologe pri govedu in človeku z najvišjo oceno identitete. Za ugotavljanje možnih genov prenašalne RNA (tRNA) smo uporabili orodje tRNAscan-SE [78] in odkrili 24 genov tRNA, od katerih je 21 predvidoma psevdogenov. Preostala zaporedja prepisov najverjetneje predstavljajo krajše čitalne okvirje, nekodirajoče RNA ali napake sekvenciranja.

<sup>1</sup>[www.transdecoder.sourceforge.net](http://www.transdecoder.sourceforge.net)

### 4.2.3 Določitev transkriptoma s pristopom *de novo*

Čeprav je strategija kartiranja na referenčni genom bolj občutljiva in lahko z njo ugotovljamo prepise z nižjim izražanjem ter celo odkrivamo nove prepise, je ta način odvisen od kakovosti zaporedja referenčnega genoma in so možne tudi napake (npr. odčitki, ki zajemajo večje introne so lahko izpuščeni). Zato smo v programu CLC Genomics Workbench uporabili program za določitev zaporedij prepisov s pristopom *de novo*, ki ne izkorišča zaporedij referenčnih genomov ali anotacij in na ta način odkrili 267 498 sosesk s povprečno dolžino 197 bp, sestavljenih iz odčтков, ki so se v vsakem vzorcu pojavili vsaj dvakrat. Najdaljša dolžina soseske je znašala 14 605 bp, dolžina znatnega števila sosesk pa je presegala 2 kb (tabela 4.2). V povprečju je bila soseska sestavljena iz 6,4 odčтков (SD=77,2). Iz odčтков, ki so se v vsakem vzorcu pojavili vsaj desetkrat, smo sestavili 49 815 sosesk. Nato smo jih funkcijsko anotirali v programu Blast2GO [79] ter jih uporabili kot referenčni transkriptom v nadaljnjih analizah. Pri tem smo uporabili algoritem blastx, v primeru, ko ni bilo zadetkov za dano sosesko, pa smo poglani še algoritem blastn. Nanj smo kartirali zaporedja z algoritmom programa CLC Bio. Zaporedja, ki niso kartirala na referenčni genom goveda (Btau4.0), smo prav tako združili v soseske s pristopom *de novo*. Iz odčтков, ki so bili prisotni v vsakem od vzorcev vsaj dvakrat, smo sestavili 91 tisoč sosesk. Iz odčтков, ki so bili prisotni v vsakem od vzorcev vsaj desetkrat, pa smo sestavili 86 tisoč sosesk. Te soseske morebiti predstavljajo prepise genov, ki so specifični za transkriptom kože ali pa prepise, ki v genomu goveda Btau4.0 še niso anotirani. Kar 63 245 (73,5%) sosesk se ujema z referenčnim transkriptomom kože. Kljub temu pa se 23 tisoč sosesk ne ujema z nobeno od referenčnih sosesk, kar je verjetno posledica nizke globine sekvenciranja in algoritmsko težavnega razlikovanja izooblik in podobnih prepisov s sekvenciranjem kratkih zaporedij samo iz ene od 5' in 3' smeri (angl. single-end). S pristopom *de novo* lahko odkrivamo prepise kakršnegakoli izvora, prav tako pa ta način ni odvisen od napovedovanja spojitenih mest. Tako smo s tem pristopom poleg prepisov kože in mikoplazme odkrili tudi ribosomalno RNA ter prepise drugih organizmov, npr. gliv in bakterij, ki so bile očitno prisotne v okuženih vzorcih, in jih samo s pristopom kartiranja na referenčne genome ne bi odkrili. Primer je na celično steno vezana hidrolaza (angl. cell wall-associated hydrolase), ki glede na homologijo najverjetneje izvira iz bakterije *Staphylococcus aureus* ali drugega mikroorganizma, ki je prestal izpostavljenost penicilinskim antibiotikom. Vendar pa je ta strategija občutljiva na napake sekvenciranja

ter je nagnjena k združevanju odčitkov podobnih prepisov v eno solesko. Z uporabo obeh strategij lahko premostimo nekatere od pomanjkljivosti posameznih pristopov in pridobimo celovitejši vpogled v dinamiko transkriptoma.

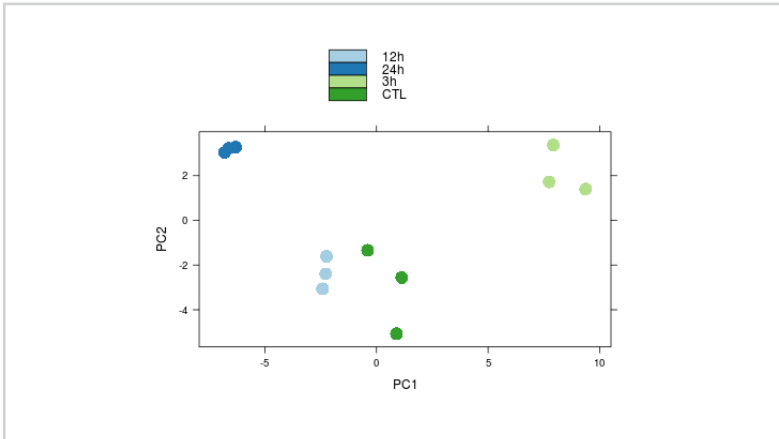
Tabela 4.2: Statistika zaporedij prepisov, določenih s pristopom *de novo*.

Število solesk	267 498
Pokritost (odčitkov/solesko)	
Povprečje	6,3
Standardni odklon	77,2
Mediana	3
Minimum	1
Maksimum	4 332
Dolžina (bp)	
Povprečje	196,5
Standardni odklon	380,0
Mediana	93
Minimum	25
Maksimum	14 605
N50	345

#### 4.2.4 Diferencialno izražanje genov

Ker je znatno število odčitkov kartiralo v območjih genoma kože, kjer ni anotacij, smo za ugotavljanje izražanja genov in nadaljnje analize uporabili le genom goveda. Analiza glavnih komponent (PCA) kaže, da se ponovitve združujejo relativno blizu skupaj in torej tri ponovitve za namene splošnega opisa raziskovanih pojavov zadoščajo. Raznolikost med tehničnimi ponovitvami je najvišja za neokužene vzorce in se s časom po okužbi manjša. Najbolj podobno izražanje imajo neokuženi vzorci in vzorci 12h po okužbi, vzorci preostalih dveh časovnih točk pa so bolj odmaknjeni (slika 4.3). Vzrok za večjo podobnost izražanja genov neokuženih celic z izražanjem genov 12h po okužbi ni jasen. Možno je, da 12h okužbe predstavlja vmesno točko med odzivom genov za nenadni in hiter odgovor na okužbo ter odzivom genov, ki na okužbo odgovorijo kasneje.

Diferencialno izražanje genov smo analizirali z uporabo različnih programskih pa-



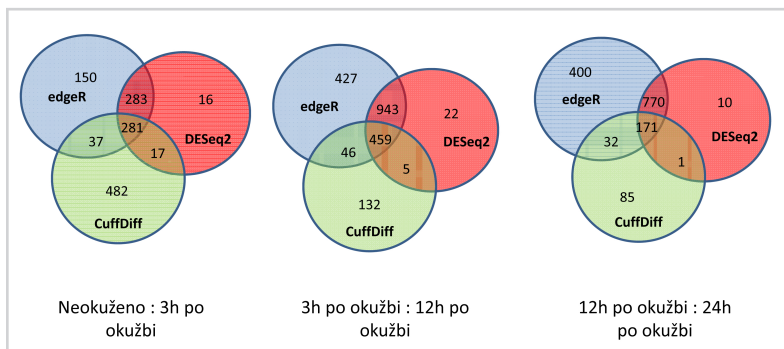
Slika 4.3

Analiza glavnih komponent za neokužene vzorce (CTL), vzorce 3h po okužbi (3h), 12h po okužbi (12h) in 24h po okužbi (24h).

ketov. Raven mRNA smo določili v smislu števila odčitkov na gen s programoma DESeq2 [80] in edgeR [81] in v smislu števila odčitkov na 1000 bp eksonskega modela na milijon kartiranih odčitkov (angl. Reads Per Kilobase of exon model per Million mapped reads, RPKM) s programoma Cufflinks in CLC Genomics Workbench. Mera genskega izražanja RPKM upošteva dolžino gena/prepisa in omogoča bolj natančne primerjave ravni izražanja preko genov/prepisov znotraj posameznega vzorca. Programa DESeq2 in edgeR zahtevata kot vhodni podatek število odčitkov na gen, ki smo ga za vsak vzorec izračunali s programom Htseq-count<sup>2</sup> z vhodnima datotekama poravnave (v obliki zapisa BAM) in anotacije genoma. S programom CLC Genomics Workbench smo statistično razliko v ravni izražanja ocenili z Baggerlyevim testom [82], ki podatke prilagaja beta-binomski porazdelitvi. Kot diferencialno izražene gene smo opredelili gene s P-vrednostjo, nižjo od 0,05, in dvakratno razliko v izražanju. Na takšen način smo odkrili 28 diferencialno izraženih genov med neokuženimi vzorci in vzorci 3h po okužbi, 47 diferencialno izraženih genov med vzorci 3h in 12h po okužbi ter 16 diferencialno izraženih genov med vzorci 12h in 24h po okužbi. DESeq2, edgeR in CuffDiff ocenjujejo statistično značilnost diferencialnega izražanja s prilaganjem podatkov negativni binomski porazdelitvi, za katero je bilo že dokazano, da omejuje statistične napake prvega tipa in predstavlja izboljšavo glede na teste razmerja

<sup>2</sup>[www-huber.embl.de/users/anders/HTSeq/](http://www-huber.embl.de/users/anders/HTSeq/)

verjetij (angl. likelihood ratio). Značilne razlike v izražanju smo ugotovili, ko je bila  $q$ -vrednost glede na Benjamini-Hochbergov postopek FDR (angl. false discovery rate) nižja od 0,05. Če primerjamo metode med seboj, so edgeR, DESeq2 in CuffDiff v tem zaporedju določili 751, 597 in 817 diferencialno izraženih genov med neokuženimi vzorci in vzorci 3h po okužbi; 1875, 1429 in 642 diferencialno izraženih genov med vzorci 3h in 12h po okužbi ter 1373, 952 in 289 diferencialno izraženih genov med vzorci 12h in 24h po okužbi (slika 4.4). Smiselno število genov je bilo odkrito kot diferencialno izraženo z vsemi tremi programi, večja skladnost pa je opazna med paketoma edgeR in DESeq2.



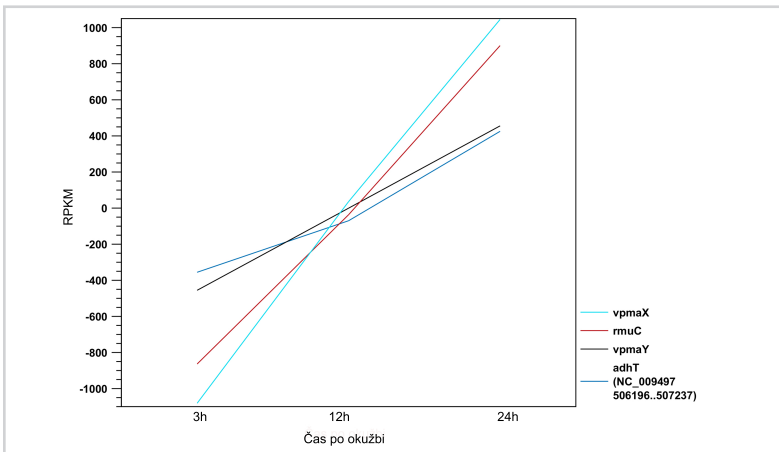
Slika 4.4

Vennovi diagrami diferencialno izraženih genov, kot smo jih odkrili s tremi različnimi statističnimi metodami.

Diferencialno izražanje genov smo testirali tudi za zaporedja, kartirana na transkriptom, določen s pristopom *de novo*. V tem primeru smo število kartiranih zaporedij normalizirali glede na celotno število kartiranih zaporedij ter glede na dolžino soeske na podoben način, kot to določa RPKM. Za diferencialno izražene gene smo opredelili prag vsaj dvakratne razlike. S tem pristopom smo odkrili 62 diferencialno izraženih genov kože med neokuženimi vzorci in 3h okuženimi vzorci, 88 diferencialno izraženih genov med vzorci 3h in 12h po okužbi in 66 diferencialno izraženih genov med vzorci 12h in 24h po okužbi. Prav tako smo odkrili 13 in 2 diferencialno izražena mikoplazemska gena med vzorci 3h in 12h po okužbi in med vzorci 12h in 24h po okužbi v tem zaporedju.

Čeprav je bil poskus zasnovan z namenom analize evkariontskega transkriptoma, smo kartirali zaporedja, ki jih nismo uspeli poravnati z zaporedjem genoma goveda, tudi na genom *M. agalactiae*, prav tako pa smo gene mikoplazme odkrili s pristopom

*de novo*. Odkrili smo značilno povišano izražanje genov *vpmaX*, *vpmaY*, *rmucC* in *adhT* 3h, 12h in 24h po okužbi (slika 4.5). Lokus *Vpma* kodira družino variabilnih površinskih lipoproteinov. Pomemben element izogibanja imunskemu sistemu gostitelja je namreč ravno raznolikost površinskih antigenov *M. agalactiae* in zanjo je odgovorna rekombinacija, specifična za zaporedja (angl. site-specific recombination) znotraj tega lokusa. Rekombinacijski protein *rmucC* je nukleaza in ima najverjetneje vlogo v rekombinaciji DNA. Povišano izražanje gena alkoholne dehidrogenaze (*adhT*) prav tako namiguje na povezavo s patogenostjo *M. agalactiae*. Raziskave kažejo, da je alkoholna dehidrogenaza ključna za oksidacijo izopropanola v aceton, vendar pa ni jasno, kakšen je pomen te aktivnosti za patogenost bakterije [83].



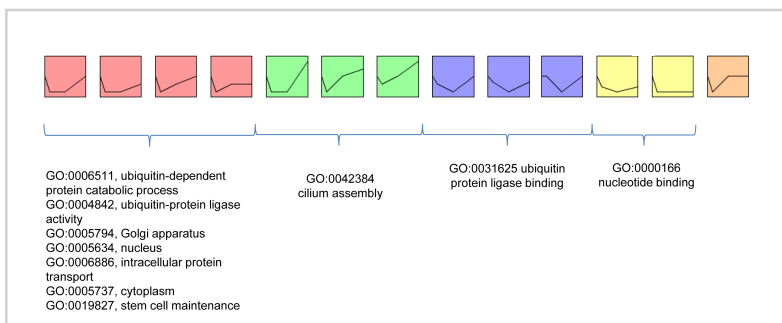
Slika 4.5

Časovno odvisno izražanje prepisov nekaterih možnih virulentnih dejavnikov mikoplazme.

Ti rezultati kažejo na pomembne virulentne dejavnike mikoplazme, kljub temu pa ne predstavljajo celotne slike prepisovanja RNA bakterije. Westermann in sodelavci [84] ocenjujejo, da je razmerje bakterijske in evkarionske RNA v vzorcih okuženih celic približno enako razmerju 1 proti 20. Tako bo za bolj celovit vpogled v proces okužbe na strani bakterij nujno potrebna večja globina sekvenciranja brez selekcije za specifične prepise RNA.

#### 4.2.5 Analiza funkcij diferencialno izraženih genov

Časovne vzorce izražanja posameznih prepisov smo razvrstili in grafično predstavili s programom STEM (Short Time series Expression Miner) v 1.3.8 [85]. Za vhodne podatke smo uporabili normalizirana povprečna števila zaporedij, kot jih podaja paket edgeR. Podatki smo prilagodili 50 možnim modelom časovno odvisnega izražanja in združili v pet različnih sklopov vzorcev časovno odvisnega izražanja. Program STEM nato izračuna statistično značilnost vsakega sklopa glede na razmerje števila genov, dodeljenih danemu profilu, proti številu pričakovanih genov in določi genske ontologije, ki so v večji meri zastopane znotraj posameznih sklopov. Odkrili smo pet glavnih sklopov profilov časovno odvisnega izražanja (slika 4.6). Prvi sklop vključuje 3492 genov z ostrim padcem izražanja 3h po okužbi in postopnim porastom izražanja po tej časovni točki. Ti geni so obogateni za od ubikvitina odvisne procese razgradnje, aktivnost ubikvitinske ligaze, znotrajcelični prenos beljakovin in vzdrževanje matičnih celic. Drugi sklop vključuje 1465 genov, ki se jim 3h po okužbi izražanje močno zniža in so obogateni za oblikovanje migetalk. Rezultat je smiseln, saj motnje v delovanju migetalk lahko upočasnijo podaljševanje mlečnih vodov in lobularno-alveolarni razvoj med nastajanjem in izločanjem mleka [86]. V tretjem sklopu je 945 genov, ki kodirajo beljakovine, sposobne vezave na ubikvitinsko ligazo. Njihovo izražanje se niža vse do 12h po okužbi, nato pa začne postopno naraščati. V četrtem sklopu je 922 genov, ki kodirajo domene za vezavo nukleotidov.

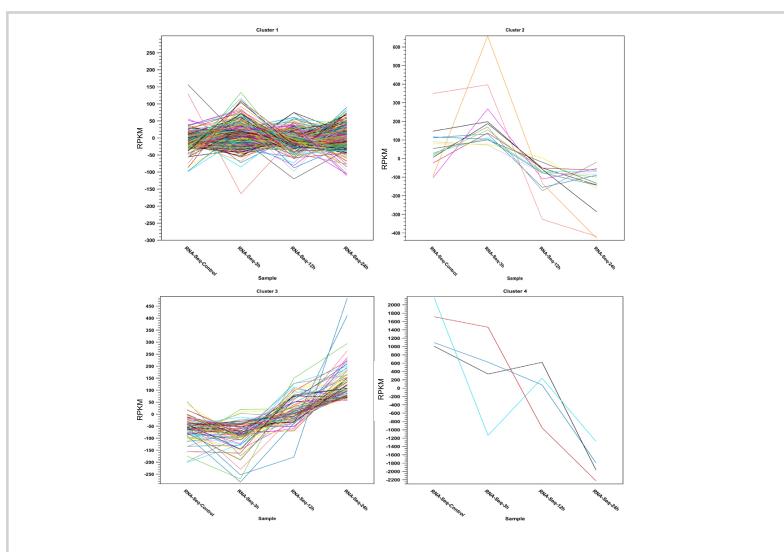


Slika 4.6

Časovni vzorci prepisovane RNA.

Gene smo glede na podobne vzorce izražanja razdelili v skupine še s postopkom *k*-tih povprečij (angl. *k*-means) v programu CLC Genomics Workbench. Glede na izražanje smo jih razvrstili v štiri skupine (slika 4.7). Največ genov v poskusu ni bistve-

no spremenilo izražanja. V drugi skupini so geni, ki imajo povišano izražanje 3h po okužbi, nato pa njihovo izražanje pade. To so geni, ki kodirajo beljakovine citoskeleta, beljakovine toplotnega šoka in druge beljakovine, pomembne za odziv organizma ob stresu. V tretji skupini so geni, ki se jim je med okužbo izražanje (skoraj eksponentno) poviševalo. To so geni, pomembni za oblikovanje ribosomov, podaljševanje in prevajanje beljakovin, geni za beljakovine ubikvitinske skupine, kalcij-vezavne beljakovine za imunski odziv in beljakovine, ki uravnavajo homeostazo in celično smrt. V četrti skupini so geni za nekatere keratine, pomembne kot prva fizična prepreka okužbe, na kateri je običajno sloj bakteriocidnih snovi. Keratini kažejo težnjo nižanja izražanja med okužbo z mikoplazmo.



*Slika 4.7*

Skupine genov glede na algoritem k-tih povprečij.

Diferencialno izražene gene smo analizirali glede na gensko ontologijo (GO). Analizo za ugotavljanje obogatenosti s pripisi GO in biološkimi potmi iz podatkovne zbirke KEGG (Kyoto Encyclopedia of Genes and Genomes) smo izvedli s paketom Goseq [87], ki popravlja izbor genov glede na njihovo dolžino, saj ta vpliva na varianco ocen genskega izražanja. Analizo smo opravili za diferencialno izražene gene, ki smo jih odkrili s paketoma edgeR in DESeq2. Rezultati obeh vhodnih naborov podatkov so

bili primerljivi, zato v nadaljevanju predstavljamo biološke poti diferencialno izraženih genov, odkritih s paketom edgeR. V spodnjih tabelah predstavljamo biološke poti s P-vrednostjo nižjo od 0,05 (tabela 4.3, tabela 4.4, tabela 4.5).

Tabela 4.3: Biološke poti s spremenjenim uravnavanjem med neokuženimi celicami in celicami 3h po okužbi.

N(D): število diferencialno izraženih genov v poti KEGG, N(I): število izraženih genov v poti KEGG, N: število anotiranih genov v poti KEGG.

Pot KEGG	P-vrednost	FDR	N(D)	N(I)	N
Cell cycle	1.57E-05	3.51E-03	21	118	124
Ribosome biogenesis in eukaryotes	4.62E-04	4.98E-02	13	70	85
p53 signaling pathway	6.66E-04	4.98E-02	11	58	68
Complement and coagulation cascades	1.98E-03	9.55E-02	6	20	68
TGF-beta signaling pathway	2.13E-03	9.55E-02	11	60	80
Steroid biosynthesis	3.62E-03	1.35E-01	5	16	20
Amyotrophic lateral sclerosis (ALS)	2.31E-02	6.75E-01	6	37	51
Small cell lung cancer	2.51E-02	6.75E-01	10	70	86
Leukocyte transendothelial migration	3.09E-02	6.75E-01	9	63	118
Bile secretion6	3.28E-02	6.75E-01	6	32	72
Phagosome	3.31E-02	6.75E-01	11	92	158

Tabela 4.4: Biološke poti s spremenjenim uravnavanjem med celicami 3h in 12h po okužbi.

N(D): število diferencialno izraženih genov v poti KEGG, N(I): število izraženih genov v poti KEGG, N: število anotiranih genov v poti KEGG.

Pot KEGG	P-vrednost	FDR	N(D)	N(I)	N
Cell cycle	4.42E-08	9.91E-06	43	120	124

Se nadaljuje na naslednji strani...

Tabela 4.4 – nadaljevanje

Pot KEGG	P-vrednost	FDR	N(D)	N(I)	N
DNA replication	1.52E-05	1.31E-03	16	34	38
p53 signaling pathway	1.75E-05	1.31E-03	22	58	68
Pyrimidine metabolism	7.82E-04	4.12E-02	24	86	105
Ribosome biogenesis in eukaryotes	9.20E-04	4.12E-02	22	70	85
Tight junction	4.42E-03	1.35E-01	25	86	134
Circadian rhythm	4.74E-03	1.35E-01	8	19	30
Oocyte meiosis	4.80E-03	1.35E-01	25	93	110
TGF-beta signaling pathway	5.75E-03	1.43E-01	18	59	80
Wnt signaling pathway	8.93E-03	2.00E-01	26	101	139
Phagosome	1.25E-02	2.55E-01	23	91	158
Leukocyte transendothelial migration	1.39E-02	2.60E-01	18	62	118
Bladder cancer	1.53E-02	2.63E-01	10	31	38
Bacterial invasion of epithelial cells	2.11E-02	3.16E-01	15	53	76
Homologous recombination	2.13E-02	3.16E-01	9	26	28
MAPK signaling pathway	2.26E-02	3.16E-01	39	172	259
Hematopoietic cell lineage	2.41E-02	3.17E-01	8	23	88
Mismatch repair	2.87E-02	3.57E-01	7	20	23
RNA polymerase	3.29E-02	3.88E-01	8	27	32
Gap junction	4.58E-02	5.08E-01	15	58	89
Non-homologous end-joining	4.76E-02	5.08E-01	5	12	13

Tabela 4.5: Biološke poti s spremenjenim uravnavanjem med celicami 12h in 24h po okužbi.

N(D): število diferencialno izraženih genov v poti KEGG, N(I): število izraženih genov v poti KEGG, N: število anotiranih genov v poti KEGG.

Pot KEGG	P-vrednost	FDR	N(D)	N(I)	N
Spliceosome	9.89E-13	2.01E-10	42	118	131
DNA replication	1.79E-12	2.01E-10	21	34	36
RNA transport	1.45E-09	1.08E-07	40	130	164
Ribosome biogenesis in eukaryotes	1.67E-08	9.37E-07	26	70	85
Steroid biosynthesis	8.62E-08	3.86E-06	11	16	20
Cell cycle	7.32E-07	2.72E-05	33	120	124
Aminoacyl-tRNA biosynthesis	8.49E-07	2.72E-05	16	37	66
Proteasome	1.22E-05	3.41E-04	15	42	44
Mismatch repair	1.97E-05	4.90E-04	10	20	23
Base excision repair	3.25E-04	7.28E-03	11	31	33
Taste transduction	3.79E-04	7.72E-03	6	10	53
p53 signaling pathway	6.10E-04	1.14E-02	16	60	68
Terpenoid backbone biosynthesis	1.44E-03	2.48E-02	6	13	21
Nucleotide excision repair	1.98E-03	3.17E-02	12	43	47
Valine, leucine and isoleucine degradation	2.82E-03	4.21E-02	10	34	44
Chagas disease (American trypanosomiasis)	3.92E-03	5.49E-02	16	69	104
Chemokine signaling pathway	4.57E-03	6.02E-02	19	89	189
Gap junction	5.45E-03	6.78E-02	14	58	89
Ribosome	6.01E-03	6.87E-02	17	89	135
Alanine, aspartate and glutamate metabolism	6.26E-03	6.87E-02	7	21	32
Nitrogen metabolism	6.44E-03	6.87E-02	5	12	17
Phagosome	8.97E-03	9.13E-02	19	95	158
GnRH signaling pathway	9.40E-03	9.15E-02	14	61	92

Se nadaljuje na naslednji strani...

Tabela 4.5 – nadaljevanje

Pot KEGG	P-vrednost	FDR	N(D)	N(I)	N
Antigen processing and presentation	1.08E-02	9.89E-02	9	35	82
African trypanosomiasis	1.14E-02	9.89E-02	5	13	34
Pyrimidine metabolism	1.15E-02	9.89E-02	17	86	105
Protein processing in endoplasmic reticulum	1.39E-02	1.15E-01	25	143	167
Rheumatoid arthritis	1.58E-02	1.22E-01	9	37	93
Metabolic pathways	1.59E-02	1.22E-01	109	814	1180
Osteoclast differentiation	1.83E-02	1.32E-01	15	75	132
Aldosterone-regulated sodium reabsorption	1.83E-02	1.32E-01	7	25	39
Selenocompound metabolism	1.94E-02	1.32E-01	5	15	17
Long-term depression	1.95E-02	1.32E-01	9	36	60
Homologous recombination	2.39E-02	1.58E-01	7	26	28
Cysteine and methionine metabolism	2.52E-02	1.61E-01	7	27	34
NOD-like receptor signaling pathway	2.65E-02	1.65E-01	9	39	57
mRNA surveillance pathway	2.83E-02	1.72E-01	13	64	91
Synthesis and degradation of ketone bodies	3.13E-02	1.82E-01	3	7	9
Progesterone-mediated oocyte maturation	3.24E-02	1.82E-01	13	66	86
Pathways in cancer	3.29E-02	1.82E-01	36	229	327
Purine metabolism	3.34E-02	1.82E-01	20	117	171
Leishmaniasis	3.68E-02	1.93E-01	9	41	76
Vascular smooth muscle contraction	3.76E-02	1.93E-01	13	66	131
ECM-receptor interaction	3.79E-02	1.93E-01	11	48	86
Butanoate metabolism	4.21E-02	2.10E-01	5	18	26

Se nadaljuje na naslednji strani...

Tabela 4.5 – nadaljevanje

Pot KEGG	P-vrednost	FDR	N(D)	N(I)	N
Propanoate metabolism	4.54E-02	2.19E-01	6	24	32
Salivary secretion	4.60E-02	2.19E-01	9	41	90

Tri ure po okužbi je bilo moteno uravnavanje 11 bioloških poti, med 3h in 12h po okužbi je bilo moteno uravnavanje 21 poti in med 12h in 24h po okužbi je bilo moteno uravnavanje 47 bioloških poti. Če za prag statistične značilnosti izberemo vrednost FDR, ki je nižja od 0,05, lahko potrdimo spremembe v uravnavanju poti celičnega cikla, oblikovanja ribosomov, signalne poti beljakovine p53 in podvojevanja DNA, po 24 h okužbe pa tudi poti spajalnega telesca (angl. spliceosome), prenosa RNA, proteasoma, biosinteze steroidov in mehanizmov popravljanja DNA. Poti celičnega cikla, signalna pot beljakovine p53, pot oblikovanja ribosomov in fagosoma so bile motene v vseh testiranih časovnih presledkih po okužbi. Številne raziskave so pokazale, da kronične okužbe celičnih kultur z mikoplazmo lahko povzročijo gensko nestabilnost in maligne spremembe celic [88–90]. Okužbe mikoplazem so pogoste tudi v tkivnih vzorcih predrakavih sprememb ter še bolj v samih malignih tkivih (kot podrobneje opisano v preglednih člankih [91] in [92]). V nasprotju z večino raziskav, v katerih je bila onkogeno narava mikoplazemske okužbe testirana v humanih in mišjih celičnih linijah, ki so nagnjene k spontanemu pojavu mutacij, smo v tej raziskavi odkrili onkogeno zmožnost *M. agalactiae* celo v celični kulturi prve pasaže pri živalski vrsti z redko dokumentiranimi tumorji mlečne žleze. Poleg tega smo moteno uravnavanje celičnega cikla opazili že 3h po okužbi, kar kaže na to, da se patogenost *M. agalactiae* ne izrazi le preko kronične okužbe in fenotipskih sprememb celic, temveč se lahko hitro sproži preko inhibicije aktivnosti beljakovine p53. Kljub temu je postopno kopičenje obogatenosti za podvojevanje DNA, popravljanje neujemanja, homologno rekombinacijo in podobne biološke poti razvidno že iz časovnega okvirja 24h okužbe. Že 3h po okužbi so bile glede na P-vrednost značilne tudi poti bolezni, kot sta npr. amiotrofična lateralna skleroza in drobnocelični pljučni rak. Čeprav sta rezultata lažno pozitivna ter imata vrednost FDR nad 0,05, je zanimivo, da sta to nekatere od bolezni, za katere se domneva, da pri njihovem pojavu igrajo vlogo tudi mikoplazme [93, 94]. Moten nastanek ribosomov je prav tako povezan s tumorigenostjo in napredovanjem tumorjev, saj povzroča spremembe v ribosomalnem nadzoru prevajanja beljakovin v

rakavih celicah [95]. Motnje te poti, skupaj s potmi proteasoma, spajalnega telesca in drugih poti procesiranja RNA in beljakovin, so v tej raziskavi med okužbo postajale vedno bolj izrazite. Obogatnost za pot fagosoma med celotnim poskusom kaže na fagocitno delovanje celic mlečne žleze *in vitro*. Glede na druge značilne biološke poti je bila ta aktivirana s sistemom komplementa, kemokini in diferenciacijo levkocitov iz krvotvorne celične linije. Kljub temu ni popolnoma jasno, ali gostitelj s fagocitozo učinkovito odstranjuje patogeni mikroorganizem, ali pa *M. agalactiae* uporablja ta znotrajcelični mehanizem kot strategijo za izogibanje imunskemu odzivu na podoben način, kot je bilo to potrjeno za vztrajno ohranjanje bakterije *Mycoplasma bovis* v pljučih okuženih telet [96]. V skladu s tem je bil v eksperimentalno vzpostavljeni kužni agalaksiji antigen *M. agalactiae* imunohistokemično zaznan tudi v citoplazmi degeneriranih epitelnih celic mlečne žleze [64]. V prid temu priča še obogatnost za poti bakterijskega vdora v epitelne celice in obogatnost za poti znotrajceličnih parazitskih bolezni, kot sta tripanosomioza in leishmanioza. Podatki o prepisovanju RNA 3h proti 12h po okužbi kažejo na možno vlogo signalnih poti TGF-beta, Wnt in MAPK pri odzivu na okužbo z *M. agalactiae*. Nadalje obogatnost za signaliziranje preko NLR (angl. NOD-like) receptorjev 12h proti 24h po okužbi kaže, da so bili mikoplazemski ligandi prepoznani znotrajcelično. Obogatnost za procesiranje in predstavljanje antigenov dokazuje sposobnost aktivacije ne samo prirojenega, ampak tudi pridobljenega imunskega sistema. Prekinjeni presledkovni stiki in tesni stiki, kot tudi medsebojno delovanje med zunajceličnim matriksom in celičnimi receptorji 12h in 24h po okužbi, so pomemben del vnetnega procesa. Medcelična prestopanja krvnih telesc (paracelularna diapedeza) in iztekanje imunskih snovi preko epitelnih celic spremenijo zgradbo in poškodujejo epitelne celice. Sprememba v prepustnosti tesnih stikov je tudi najbolj pogost vzrok za neravnovesje elektrolitov, na katerega kaže tudi z aldosteronom uravnavana pot prehajanja natrija preko celične membrane. Pot krčenja gladkega mišičevja priča o motenem mehanizmu iztiska mleka mioepitelnih celic. Spremenjeno je bilo uravnavanje tudi drugih poti, pomembnih za vsebnost mlečnih maščob in lipidni metabolizem, kot so npr. biosinteza steroidov in terpenoidov, sinteza in razgradnja ketonskih telesc ter metabolizem butanoata in propanoata. Razgradnjo in metabolizem mnogih amino kislin smo zaznali po 12h okužbe in ti bi lahko bili posledica proteolitske aktivnosti mikrobnih organizmov in fagocitov. Vendar pa je bilo prepisovanje genov *CSN2*, *LGB*, *LALBA*, *LTF*, *PLG* in drugih podobnih genov pod zaznavno mejo že v neokuženih celicah in tako nismo mogli določiti nedvoumne učinka okužbe z

*M. agalactiae* na sestavo mlečnih beljakovin. Zanimivo je, da so Ogorevc in sodelavci [65] z uporabo PCR z obratnim prepisom (qRT-PCR) zaznali prepisovanje gena za beta-kazein v drugi pasaži celične kulture. Tako ni jasno, ali je za to razliko v detekciji odgovorna tehnika ugotavljanja izražanja ali dejanska sprememba v izražanju gena med prvo in drugo pasažo celic.

### 4.3 Diskusija in prihodnji izzivi

V okviru tega dela smo odkrili dinamične spremembe transkriptoma v organizmih gostitelja in bakterije v celičnem modelu kužne agalaksije, nakazali možne virulentne dejavnike *M. agalactiae*, odkrili značilne biološke poti z motenim uravnavanjem v celičah mlečne žleze gostitelja in odkrili veliko zbirko podatkov o prepisovanju RNA, ki se jo lahko nadalje uporabi za anotacijo genoma kože. V nadaljevanju pa naj se dotaknemo nekaterih točk poskusa, ki dajejo prostor nadaljnjim raziskavam ter izboljšavam pri tovrstnih analizah transkriptoma.

Gene, ki smo jih odkrili kot diferencialno izražene z metodo sekvenciranja RNA, bi bilo v nadaljevanju raziskav nujno ovrednotiti in potrditi še z metodo kvantitativnega PCR z obratnim prepisom (qRT-PCR). Učinki kužne agalaksije *in vivo* so odvisni od izrazitosti okužbe ter sekundarnih fizioloških in sistemskih sprememb organizma, ki jih ni mogoče preučevati z metodami *in vitro*. Raziskave tudi kažejo, da je transkriptom mlečne žleze okužene *in vivo* bistveno drugačen od transkriptoma eksperimentalno okužene celične kulture [97], saj se eksperimentalna in naravna pot okužbe razlikujeta. Slednja namreč povzroči poleg lokalnih tudi sistemske vplive na organizem. Tako bi v prihodnjih raziskavah zasnovo poskusa lahko izboljšali z uporabo laserskega mikroskopskega razreza (mikrodisekcije) posameznih celičnih tipov iz mlečne žleze okužene *in vivo*, ali pa celo z uporabo vzorčenja mleka in preučili odzive prepisovanja RNA v različnih celičnih populacijah gostitelja. Predvsem bi bilo pomembno preučiti odziv celic, ki izražajo mlečne beljakovine, kar v tej raziskavi ni bilo mogoče. Še več, ker je okužba izredno dinamičen proces ter ker niso niti celične kulture niti tkiva homogena, bi bilo za razkrivanje odzivov celic različnih tipov najbolje transkriptom preučevati na ravni posameznih celic. Čeprav je bil poskus namenoma zasnovan s čim manjšim številom spremenljivk, brez testiranja vnaprej zastavljenih hipotez, bi robustnost intrinzične spremenljivosti izražanja in ponovljivost rezultatov še izboljšali z biološkimi ponovitvami. Ker je število poti s spremenjenim uravnavanjem med poskusom vztrajno naraščalo, bi bilo v prihodnje smiselno opraviti raziskave v širšem časovnem okviru.

Ker je analiza razkrila pomembno vlogo spajalnega telesca, bi bilo zanimivo podobno raziskavo opraviti tudi s sekvenciranjem iz obeh (5' in 3') smeri (angl. paired-end), s katerim lahko določimo zaporedja obeh koncev fragmentov RNA in tako omogočimo boljše odkrivanje alternativnega spajanja eksonov. Glede na to, da naša in mnoge druge raziskave kažejo onkogeno zmožnost mikoplazme, bi lahko prihodnje raziskave z uporabo visokozmogljivostnega sekvenciranja DNA razkrile tudi specifične kromosomske in genske spremembe, ki so vzrok motnjam celičnega cikla. Za hkratno opazovanje časovne dinamike izražanja bakterije in celic gostitelja bi bilo koristno raziskavo vnaprej zasnovati brez selekcije za specifične prepise, ki vodi v delno izgubo celokupnega transkriptoma. S tem bi omogočili ugotavljanje izražanja večjega števila genov mikoplazme in posledično še bolj poglobljeno razumevanje medsebojnih vplivov med evkariontskim gostiteljem in prokariontskim patogenom.

Čeprav ima sekvenciranje RNA mnogo prednosti pred ugotavljanjem izražanja genov z mikromrežami in nima napak, ki so povezane s hibridizacijo prepisov, pa je še vedno prisoten računski izziv hkratnega testiranja več hipotez. Ker z metodo sekvenciranja RNA opazujemo izražanje velikega števila genov, je možnost, da bodo nekateri od teh izraženi diferencialno zgolj po naključju, toliko večja. Z iskanjem bioloških mehanizmov izmed množice kandidatnih bioloških poti postane ta statistična pomanjkljivost še bolj resna. Analiza bioloških poti, v katere so vključeni geni s spremenjenim izražanjem, je v našem delu poleg biološko smiselnih dala tudi biološke poti, ki so najverjetneje lažno pozitivne, kot so npr. signalne poti gonadoliberina, diferenciacije osteoklastov, dozorevanja in mejoze oocit, depresije in druge poti. V to analizo so bili namreč vključeni vsi diferencialno izraženi geni, ki so imeli vrednost FDR nižjo od 0,05. To pomeni, da smo hkrati z večjo občutljivostjo testiranja, ki jo omogoča FDR, v analizo bioloških poti vključili tudi 5% genov, ki so bili najverjetneje lažno pozitivni. Na raven napačnih rezultatov vplivajo tudi vsi predhodni koraki analize, npr. kartiranje odčitkov kože na genom goveda in uporaba funkcije, ki popravlja dolžine genov goveda pri analizi obogatenosti poti, pri čemer smo se torej zanašali na podobnost ortolognih genov obeh živalskih vrst. Kljub temu se je večina odkritih bioloških poti diferencialno izraženih genov, kot smo jih odkrili s paketom edgeR, skladala z biološkimi potmi, ki smo jih odkrili na osnovi diferencialno izraženih genov, odkritih s paketom DESeq2. Od zgoraj navedenih, verjetno lažno pozitivnih poti, je bila z uporabo podatkov izražanja genov, ugotovljenih s paketom DESeq2, ponovno odkrita le pot diferenciacije osteoklastov. V preučevanem poskusu ni bilo spremenjeno izražanje

genov, ki so specifični za osteoklaste (npr. geni *CTSK*, *TRAP*, *CTR*, *INTB3*), vendar pa so bili diferencialno izraženi nekateri geni celičnega cikla, signalne poti MAPK in imunskih receptorjev, ki so vpleteni tudi v diferenciacijo osteoklastov in tako vsaj deloma pojasnjujejo takšen rezultat. Razvoj algoritmov, ki bi imeli čim večjo občutljivost za resnično pozitivne rezultate, hkrati pa bi omejevali lažno pozitivne napake, bo obvezen za uspeh analiz obogatenosti genov, hkrati pa bo tudi v naprej za natančnost rezultatov pomembno zadostno število bioloških in tehničnih ponovitev ter potrditev rezultatov z dodatnimi eksperimentalnimi tehnikami ugotavljanja izražanja genov.

## *Zaključek*

V doktorski nalogi se posvečamo genetiki kompleksnih lastnosti pri domačih živalih in jih analiziramo z različnimi računskimi pristopi. Soočamo se z izzivi zasnove eksperimentov kot tudi njihove analize, pomanjkljivimi anotacijami genomov preučevanih organizmov ter analizami funkcij genov. Kot fenotipsko najbolj raznolika vrsta je pes dober modelni organizem za raziskovanje genske pestrosti in znotrajvrstno primerjavo pasem. Pastirske pasme Zahodnega Balkana so dober primer kompleksnega vzorca genske raznolikosti, saj kažejo, kljub skupnemu geografskemu izvoru, podobnemu fenotipu in relativno pozni ločitvi v posamezne pasme, različne značilnosti populacijske dinamike. Z namenom ocene genske pestrosti in razlikovanja pasem ter odkrivanja evolucijskih mehanizmov, ki vplivajo na odnose med pasmami, smo uporabili različne analitske metode populacijske genetike. Rezultati naše raziskave kažejo, da stohastični dogodki v majhnih populacijah izrazito vplivajo na sorodstvene odnose in spreminjajo populacijsko strukturo ter s tem tudi gensko razlikovanje, prav tako pa imajo lahko fenotipske učinke, npr. pojav dednih boleznih kot posledica depresije zaradi inbreedinga.

V poglavju 3 raziskujemo gensko ozadje ektopije sečevodov, ki je najbolj pogosta razvojna napaka uhajanja seča pri mladih psih, klinični znaki bolezni pa so kot posledica nepopolne genske penetrance lahko izredno raznoliki. Povišano pojavnost smo opazili pri kraškem ovčarju. Da bi odkrili kandidatne lokuse, smo analizirali rodovnike ter opravili asociacijsko preiskavo celotnega genoma. Z genotipizacijo petih psov z ekopičnimi sečevodi ter petih psov brez kliničnih znakov na 170k SNP-čipu smo odkrili več položajskih in funkcijskih kandidatnih genov na kromosomih CFA11, CFA18 in CFA20 ter potrdili kompleksen vzorec dedovanja, kot ga je nakazala že analiza rodovnikov. Odkrivanje mendelskih neskladij, kartiranje homozigotnosti in uporaba genotipskih podatkov psov drugih pasem so dodatno nakazali možne vloge teh lokusov. Z analizo funkcij genov smo razkrili, da so vsi kandidatni geni udeleženi v bioloških poteh celičnega cikla in embrionalnega razvoja. Zaradi majhnega števila analiziranih vzorcev so rezultati zaenkrat še nedokončni vendar sugestivni.

V poglavju 4 prikazujemo rezultate določitve transkriptoma celične kulture mlečne žleze kože okužene *in vitro* z bakterijo *Mycoplasma agalactiae* z uporabo referenčnih zaporedij genomov in s pristopom brez takšne uporabe. Odkriti prepisi lahko služijo kot vir za izboljšavo strukturne anotacije genoma kože, diferencialno izražanje genov pa je razkrilo pomembne virulentne dejavnike bakterije, dinamične vzorce transkriptomov ter številne spremembe v uravnavanju bioloških poti v gostiteljevem organizmu,

Bioinformacijski analitski pristopi

65

ki zahtevajo nadaljnje raziskave.





## LITERATURA

- [1] H. G. Parker, L. V. Kim, N. B. Sutter, S. Carlson, T. D. Lorenzen, T. B. Malek, G. S. Johnson, H. B. DeFrance, E. A. Ostrander, and L. Kruglyak. Genetic structure of the purebred domestic dog. *Science*, 304:1160–1164, 2004. ISSN 1095-9203. doi: [10.1126/science.1097406](https://doi.org/10.1126/science.1097406). URL <http://www.ncbi.nlm.nih.gov/pubmed/15155949>.
- [2] G. Leroy, E. Verrier, J. C. Meriaux, and X. Rognon. Genetic diversity of dog breeds: between-breed diversity, breed assignment and conservation approaches. *Anim Genet*, 40:333–343, 2009. ISSN 1365-2052. doi: [10.1111/j.1365-2052.2008.01843.x](https://doi.org/10.1111/j.1365-2052.2008.01843.x). URL <http://www.ncbi.nlm.nih.gov/pubmed/19222436>.
- [3] A. E. Pires, I. R. Amorim, C. Ginja, M. Gomes, I. Godinho, F. Simões, M. Oom, F. Petrucci-Fonseca, J. Matos, and M. W. Bruford. Molecular structure in peripheral dog breeds: Portuguese native breeds as a case study. *Anim Genet*, 40:383–392, 2009. ISSN 1365-2052. doi: [10.1111/j.1365-2052.2009.01849.x](https://doi.org/10.1111/j.1365-2052.2009.01849.x). URL <http://www.ncbi.nlm.nih.gov/pubmed/19298456>.
- [4] H. G. Parker, A. V. Kukekova, D. T. Akey, O. Goldstein, E. F. Kirkness, K. C. Baysac, D. S. Mosher, G. D. Aguirre, G. M. Acland, and E. A. Ostrander. Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. *Genome Res*, 17:1562–1571, 2007. ISSN 1088-9051. doi: [10.1101/gr.6772807](https://doi.org/10.1101/gr.6772807). URL <http://www.ncbi.nlm.nih.gov/pubmed/17916641>.
- [5] B. M. Vonholdt, J. P. Pollinger, K. E. Lohmueller, E. Han, H. G. Parker, P. Quignon, J. D. Degenhardt, A. R. Boyko, D. A. Earl, A. Auton, A. Reynolds, K. Bryc, A. Brisbin, J. C. Knowles, D. S. Mosher, T. C. Spady, A. Elkahoulou, E. Geffen, M. Pilot, W. Jedrzejewski, C. Greco, E. Randi, D. Bannasch, A. Wilton, J. Shearman, M. Musiani, M. Cargill, P. G. Jones, Z. Qian, W. Huang, Z. L. Ding, Y. P. Zhang, C. D. Bustamante, E. A. Ostrander, J. Novembre, and R. K. Wayne. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*, 464:898–902, 2010. ISSN 1476-4687. doi: [10.1038/nature08837](https://doi.org/10.1038/nature08837). URL <http://www.ncbi.nlm.nih.gov/pubmed/20237475>.
- [6] J. Berglund, E. M. Nevalainen, A. M. Molin, M. Perloski, C. André, M. C. Zody, T. Sharpe, C. Hitte, K. Lindblad-Toh, H. Lohi, M. T. Webster, and The LUPA Consortium. Novel origins of copy number variation in the dog genome. *Genome Biol*, 13:R73, 2012. ISSN 1465-6914. doi: [10.1186/gb-2012-13-8-173](https://doi.org/10.1186/gb-2012-13-8-173). URL <http://www.ncbi.nlm.nih.gov/pubmed/22916802>.
- [7] N. Pedersen, H. Liu, G. Theilen, and B. Sacks. The effects of dog breed development on genetic diversity and the relative influences of performance and conformation breeding. *Journal of animal breeding and genetics = Zeitschrift für Tierzucht und Züchtungsbiologie*, 130(3):236–48, June 2013. ISSN 1439-0388. doi: [10.1111/jbg.12017](https://doi.org/10.1111/jbg.12017). URL <http://www.ncbi.nlm.nih.gov/pubmed/23679949>.
- [8] S. Björnerfeldt, F. Hailer, M. Nord, and C. Villà. Assortative mating and fragmentation within dog breeds. *BMC Evol Biol*, 8:28, 2008. ISSN 1471-2148. doi: [10.1186/1471-2148-8-28](https://doi.org/10.1186/1471-2148-8-28). URL <http://www.ncbi.nlm.nih.gov/pubmed/18226210>.
- [9] K. Streitberger, M. Schweizer, R. Kropatsch, G. Dekomien, O. Distl, M. S. Fischer, J. T. Epplen, and S. T. Herrwig. Rapid genetic diversification within dog breeds as evidenced by a case study on Schnauzers. *Animal genetics*, 43(5):577–86, October 2012. ISSN 1365-2052. doi: [10.1111/j.1365-2052.2011.02300.x](https://doi.org/10.1111/j.1365-2052.2011.02300.x). URL <http://www.ncbi.nlm.nih.gov/pubmed/22497658>.
- [10] M. Urosevic, P. Turina, G. Stanivukovic, and M. Jovanovic. *Balkanski pastirski psi*. Samozalozba, 2002.
- [11] B. Spoljaric. *Hrvatska kinoloska bastina*. Hrvatski kinoloski savez, 2008.
- [12] V. Krstevski. *The Macedonian Highlander – Sharplaninac*. Kumanovo, 2010.

- [13] M. Židar. *O kraševcu*. Samozaložba, Domžale, 1990.
- [14] J. P. Gutiérrez, L. J. Royo, I. Alvarez, and F. Goyache. MolKin v2.0: a computer program for genetic analysis of populations using molecular coancestry information. *J Hered*, 96:718–721, 2005. ISSN 0022-1503. doi: [10.1093/jhered/esl118](https://doi.org/10.1093/jhered/esl118). URL <http://www.ncbi.nlm.nih.gov/pubmed/16251515>.
- [15] Z. A. Szpiech, M. Jakobsson, and N. A. Rosenberg. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, 24:2498–2504, 2008. ISSN 1367-4811. doi: [10.1093/bioinformatics/btn478](https://doi.org/10.1093/bioinformatics/btn478). URL <http://www.ncbi.nlm.nih.gov/pubmed/18779233>.
- [16] A. Caballero and M. A. Toro. Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics*, 3(3):289–299, September 2002. ISSN 1572-9737. doi: [10.1023/A:1019956205473](https://doi.org/10.1023/A:1019956205473). URL <http://link.springer.com/article/10.1023/A:1019956205473>.
- [17] R. J. Petit, A. El Mousadik, and O. Pons. Identifying Populations for Conservation on the Basis of Genetic Markers. *Conservation Biology*, 12(4):844–855, July 1998. ISSN 08888892. doi: [10.1111/j.1523-1739.1998.96489.x](https://doi.org/10.1111/j.1523-1739.1998.96489.x). URL <http://doi.wiley.com/10.1111/j.1523-1739.1998.96489.x>.
- [18] A. R. Boyko, R. H. Boyko, C. M. Boyko, H. G. Parker, M. Castelhanlo, L. Corey, J. D. Degenhardt, A. Auton, M. Hedimbi, R. Kityo, E. A. Ostrander, J. Schoenebeck, R. J. Tothunter, P. Jones, and C. D. Bustamante. Complex population structure in African village dogs and its implications for inferring dog domestication history. *Proc Natl Acad Sci U S A*, 106:13903–13908, 2009. ISSN 1091-6490. doi: [10.1073/pnas.0902129106](https://doi.org/10.1073/pnas.0902129106). URL <http://www.ncbi.nlm.nih.gov/pubmed/19666600>.
- [19] L. Pereira, B. Van Asch, and A. Amorim. Standardisation of nomenclature for dog mtDNA D-loop: a prerequisite for launching a Canis familiaris database. *Forensic science international*, 141(2-3):99–108, May 2004. ISSN 0379-0738. doi: [10.1016/j.forsciint.2003.12.014](https://doi.org/10.1016/j.forsciint.2003.12.014). URL <http://www.ncbi.nlm.nih.gov/pubmed/15062947>.
- [20] L. Excoffier, G. Laval, and S. Schneider. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online*, 1:47–50, 2005. ISSN 1176-9343. URL <http://www.ncbi.nlm.nih.gov/pubmed/19325852>.
- [21] K. M. Webb and M. W. Allard. Identification of forensically informative SNPs in the domestic dog mitochondrial control region. *Journal of forensic sciences*, 54(2):289–304, March 2009. ISSN 1556-4029. doi: [10.1111/j.1556-4029.2008.00953.x](https://doi.org/10.1111/j.1556-4029.2008.00953.x). URL <http://www.ncbi.nlm.nih.gov/pubmed/19261051>.
- [22] J. P. Gutiérrez and F. Goyache. A note on EN-DOG: a computer program for analysing pedigree information. *J Anim Breed Genet*, 12:2: 172–176, 2005. ISSN 0931-2668. URL <http://www.ncbi.nlm.nih.gov/pubmed/16130468>.
- [23] D. Boichard. Pedig : a fortran package for pedigree analysis suited to large populations. In *7th World Congress on Genetics Applied to Livestock Production (Ed. by WCGALP Organizing Committee)*, pages CD–ROM no. 28–13, Montpellier, France, 2002.
- [24] K. Mäki. Population structure and genetic diversity of worldwide Nova Scotia Duck Tolling Retriever and Lancashire Heeler dog populations. *Journal of animal breeding and genetics = Zeitschrift für Tierzucht und Züchtungsbiologie*, 127(4):318–26, August 2010. ISSN 1439-0388. doi: [10.1111/j.1439-0388.2010.00851.x](https://doi.org/10.1111/j.1439-0388.2010.00851.x). URL <http://www.ncbi.nlm.nih.gov/pubmed/20646119>.
- [25] S. R. Urfér. Inbreeding and fertility in Irish Wolfhounds in Sweden: 1976 to 2007. *Acta veterinaria Scandinavica*, 51(1):21, January 2009. ISSN 1751-0147. doi: [10.1186/1751-0147-51-21](https://doi.org/10.1186/1751-0147-51-21). URL <http://www.actavetscand.com/content/51/1/21>.
- [26] I. Głażewska. Genetic diversity in Polish hounds estimated by pedigree analysis. *Livestock Science*, 113(2-3):296–301, February 2008. ISSN 18711413. doi: [10.1016/j.livsci.2007.06.012](https://doi.org/10.1016/j.livsci.2007.06.012). URL [http://www.livestockscience.com/article/S1871-1413\(07\)00398-8/abstract](http://www.livestockscience.com/article/S1871-1413(07)00398-8/abstract).
- [27] P. Bijma. Long-term genetic contributions: prediction of rates of inbreeding and genetic gain in selected populations. 2000. URL <http://www.cabdirect.org/abstracts/20000109374.html;jsessionid=ED1DC249E204CDB2AFB27F3A061104D5>.
- [28] G. Leroy, C. Danchin-Burge, I. Palthiere, R. Baumung, S. Fritz, J. C. Mériaux, and M. Gautier. An ABC estimate of pedigree error rate: application in dog, sheep and cattle breeds. *Anim Genet*, 43:309–314, 2012. ISSN 1365-2052. doi: [10.1111/j.1365-2052.2011.02253.x](https://doi.org/10.1111/j.1365-2052.2011.02253.x). URL <http://www.ncbi.nlm.nih.gov/pubmed/22486502>.
- [29] O. Langella. Populations, a free population genetic software, 2002. URL [http://www.bioinformatics.org/project/?group\\_id=84](http://www.bioinformatics.org/project/?group_id=84).
- [30] D. Bryant and V. Moulton. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol*, 21:255–265, 2004. ISSN 0737-4038. doi: [10.1093/molbev/msh018](https://doi.org/10.1093/molbev/msh018). URL <http://www.ncbi.nlm.nih.gov/pubmed/14660700>.
- [31] D. H. Huson and D. Bryant. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*, 23(2):254–67, February 2006. ISSN 0737-4038. doi: [10.1093/molbev/mj030](https://doi.org/10.1093/molbev/mj030). URL <http://mbe.oxfordjournals.org/content/23/2/254>.

- [32] J. Felsenstein. PHYLP (Phylogeny Inference Package) Version 3.6, 2005.
- [33] J. K. Pritchard, M. Stephens, and P. Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155:945–959, 2000. ISSN 0016-6731. URL <http://www.ncbi.nlm.nih.gov/pubmed/10835412>.
- [34] G. Evanno, S. Regnaut, and J. Goudet. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*, 14:2611–2620, 2005. ISSN 0962-1083. doi: 10.1111/j.1365-294X.2005.02553.x. URL <http://www.ncbi.nlm.nih.gov/pubmed/15969739>.
- [35] B. Rannala and J. L. Mountain. Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci U S A*, 94:9197–9201, 1997. ISSN 0027-8424. URL <http://www.ncbi.nlm.nih.gov/pubmed/9256459>.
- [36] S. Piry, A. Alapetite, J. M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. GENECLASS2: a software for genetic assignment and first-generation migrant detection. *J Hered*, 95:536–539, 2004. ISSN 0022-1503. doi: 10.1093/jhered/esh074. URL <http://www.ncbi.nlm.nih.gov/pubmed/15475402>.
- [37] T. Jombart, S. Devillard, and F. Balloux. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*, 11:94, 2010. ISSN 1471-2156. doi: 10.1186/1471-2156-11-94. URL <http://www.ncbi.nlm.nih.gov/pubmed/20950446>.
- [38] V. Dimitrijević. *Genotipizacija jugoslovenskog ovčarskog psa šarplaninca primenom mikrosatelitskih genetičkih markera*. PhD thesis, Univerzitet u Beogradu, 2008.
- [39] H. J. Huson, H. G. Parker, J. Runstadler, and E. A. Ostrander. A genetic dissection of breed composition and performance enhancement in the Alaskan sled dog. *BMC Genet*, 11:71, 2010. ISSN 1471-2156. doi: 10.1186/1471-2156-11-71. URL <http://www.ncbi.nlm.nih.gov/pubmed/20649949>.
- [40] D. J. Lawson, G. Hellenthal, S. Myers, and D. Falush. Inference of population structure using dense haplotype data. *PLoS genetics*, 8(1):e1002453, January 2012. ISSN 1553-7404. doi: 10.1371/journal.pgen.1002453. URL <http://dx.plos.org/10.1371/journal.pgen.1002453>.
- [41] N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, and D. Reich. Ancient admixture in human history. *Genetics*, 192(3):1065–93, November 2012. ISSN 1943-2631. doi: 10.1534/genetics.112.145037. URL <http://www.ncbi.nlm.nih.gov/pubmed/22960212>.
- [42] J. K. Pickrell and J. K. Pritchard. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS genetics*, 8(11):e1002967, January 2012. ISSN 1553-7404. doi: 10.1371/journal.pgen.1002967. URL <http://dx.plos.org/10.1371/journal.pgen.1002967>.
- [43] P. Loh, M. Lipson, N. Patterson, P. Moorjani, J. K. Pickrell, D. Reich, and B. Berger. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics*, 193(4):1233–54, April 2013. ISSN 1943-2631. doi: 10.1534/genetics.112.147330. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3606100&tool=pmcentrez&rendertype=abstract>.
- [44] J. Sutherland-Smith, R. M. Jerram, A. M. Walker, and C. G. A. Warman. Ectopic Ureters and Ureteroceles in Dogs: Presentation, cause, and diagnosis. *Compendium on Continuing Education for the Practicing Veterinarian*, 26:303–310, 2004.
- [45] H. M. Hayes. Breed associations of canine ectopic ureter: a study of 217 female cases. *Journal of Small Animal Practice*, 25(8):501–504, August 1984. ISSN 0022-4510. doi: 10.1111/j.1748-5827.1984.tb03422.x. URL <http://doi.wiley.com/10.1111/j.1748-5827.1984.tb03422.x>.
- [46] P. E. Holt and A. H. Moore. Canine ureteral ectopia: an analysis of 175 cases and comparison of surgical treatments. *The Veterinary record*, 136(14):345–9, April 1995. ISSN 0042-4900. URL <http://www.ncbi.nlm.nih.gov/pubmed/7610537>.
- [47] C. North, J. M. Kruger, P. J. Venta, J. M. Miller, D. S. Rosenstein, E. K. Randall, B. White, and S. D. Fitzgerald. Congenital ureteral ectopia in continent and incontinent-related Entlebucher mountain dogs: 13 cases (2006–2009). *Journal of Veterinary Internal Medicine / American College of Veterinary Internal Medicine*, 24(5):1055–62, 2010. ISSN 0891-6640. doi: 10.1111/j.1939-1676.2010.0562.x. URL <http://www.ncbi.nlm.nih.gov/pubmed/20666980>.
- [48] F. N. Bitterli. *Prävalenz und klinische Relevanz ectopischer Ureteren beim Entlebucher und Appenzeler Sennenhund*. Inaugural-dissertation zur erlangung der doktorwürde der vetsuisse-fakultät universität zürich, Der Vetsuisse-Fakultät Universität Zürich, 2011.
- [49] I. M. Reichler, C. Eckrich Specker, M. Hubler, A. Boos, M. Haessig, and S. Arnold. Ectopic ureters in dogs: clinical features, surgical techniques and outcome. *Veterinary surgery : VS*, 41(4):515–22, May 2012. ISSN 1532-950X. doi: 10.1111/j.1532-950X.2012.00952.x. URL <http://www.ncbi.nlm.nih.gov/pubmed/22486391>.
- [50] R. Fritsche, G. Dolf, C. Schelling, S. O. Hungerbuehler, R. Hagen, and I. M. Reichler. Inheritance of ectopic ureters in Entlebucher Mountain Dogs. *Journal of*

- Animal Breeding and Genetics*, pages n/a–n/a, September 2013. ISSN 09312668. doi: [10.1111/jbg.12055](https://doi.org/10.1111/jbg.12055). URL <http://doi.wiley.com/10.1111/jbg.12055>.
- [51] T. Fossum. Ectopic ureter. In *Small Animal Surgery*, page 1632. Elsevier Health Sciences, 2007.
- [52] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, 81(3):559–75, September 2007. ISSN 0002-9297. doi: [10.1086/519795](https://doi.org/10.1086/519795). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1950838&tool=pmcentrez&rendertype=abstract>.
- [53] J. C. Barrett, B. Fry, J. Maller, and M. J. Daly. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics (Oxford, England)*, 21(2):263–5, January 2005. ISSN 1367-4803. doi: [10.1093/bioinformatics/bth457](https://doi.org/10.1093/bioinformatics/bth457). URL <http://www.ncbi.nlm.nih.gov/pubmed/15297300>.
- [54] V. Segura, B. J. Vilhjálmsson, A. Platt, A. Korte, U. Seren, Q. Long, and M. Nordborg. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature genetics*, 44(7):825–30, July 2012. ISSN 1546-1718. doi: [10.1038/ng.2314](https://doi.org/10.1038/ng.2314). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3386481&tool=pmcentrez&rendertype=abstract>.
- [55] R. D. Thiagarajan, K. M. Georgas, B. A. Rumballe, E. Lesieur, H. S. Chiu, D. Taylor, D. T. P. Tang, S. M. Grimmond, and M. H. Little. Identification of anchor genes during kidney development defines ontological relationships, molecular subcompartments and regulatory pathways. *PLoS One*, 6(2):e17286, January 2011. ISSN 1932-6203. doi: [10.1371/journal.pone.0017286](https://doi.org/10.1371/journal.pone.0017286). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3046260&tool=pmcentrez&rendertype=abstract>.
- [56] G. Martinez, K. Georgas, G. A. Challen, B. Rumballe, M. J. Davis, D. Taylor, R. D. Teasdale, S. M. Grimmond, and M. H. Little. Definition and spatial annotation of the dynamic secretome during early kidney development. *Developmental dynamics : an official publication of the American Association of Anatomists*, 235(6):1709–19, June 2006. ISSN 1058-8388. doi: [10.1002/dvdy.20740](https://doi.org/10.1002/dvdy.20740). URL <http://www.ncbi.nlm.nih.gov/pubmed/16538671>.
- [57] J. A. Davies, C. B. Millar, E. M. Johnson, and J. Milbrandt. Neurturin: an autocrine regulator of renal collecting duct development. *Developmental genetics*, 24(3-4):284–92, January 1999. ISSN 0192-253X. doi: [10.1002/\(SICI\)1520-6408\(1999\)24:3<284::AID-DVG1>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1520-6408(1999)24:3<284::AID-DVG1>3.0.CO;2-X). URL <http://www.ncbi.nlm.nih.gov/pubmed/10932636>.
- [58] S. Ohno, S. Nakagawara, Y. Honda, and S. Nakajin. Evidence for expression of 11 $\beta$ -hydroxysteroid dehydrogenase type3 (HSD11B3/HSD11B1L) in neonatal pig testis. *Molecular and cellular biochemistry*, 381(1-2):145–56, September 2013. ISSN 1573-4919. doi: [10.1007/s11010-013-1697-4](https://doi.org/10.1007/s11010-013-1697-4). URL <http://www.ncbi.nlm.nih.gov/pubmed/23881245>.
- [59] M. A. McLoughlin and D. J. Chew. Diagnosis and surgical management of ectopic ureters. *Clinical techniques in small animal practice*, 15(1):17–24, February 2000. ISSN 1096-2867. doi: [10.1053/svms.2000.7302](https://doi.org/10.1053/svms.2000.7302). URL <http://www.ncbi.nlm.nih.gov/pubmed/10911681>.
- [60] R. Geng, C. Yuan, and Y. Chen. Exploring differentially expressed genes by RNA-Seq in cashmere goat (*Capra hircus*) skin during hair follicle development and cycling. *PLoS One*, 8(4):e62704, January 2013. ISSN 1932-6203. doi: [10.1371/journal.pone.0062704](https://doi.org/10.1371/journal.pone.0062704). URL <http://dx.plos.org/10.1371/journal.pone.0062704>.
- [61] T. Xu, X. Guo, H. Wang, F. Hao, X. Du, X. Gao, and D. Liu. Differential gene expression analysis between anagen and telogen of *Capra hircus* skin based on the de novo assembled transcriptome sequence. *Gene*, 520(1):30–38, 2013. URL <http://www.sciencedirect.com/science/article/pii/S0378111913001947>.
- [62] B. Zhu, T. Xu, J. Yuan, X. Guo, and D. Liu. Transcriptome Sequencing Reveals Differences between Primary and Secondary Hair Follicle-derived Dermal Papilla Cells of the Cashmere Goat (*Capra hircus*). *PLoS One*, 8(9):e76282, January 2013. ISSN 1932-6203. doi: [10.1371/journal.pone.0076282](https://doi.org/10.1371/journal.pone.0076282). URL <http://dx.plos.org/10.1371/journal.pone.0076282>.
- [63] Y. Dong, M. Xie, Y. Jiang, N. Xiao, X. Du, W. Zhang, G. Tossier-Klopp, J. Wang, S. Yang, J. Liang, W. Chen, J. Chen, P. Zeng, Y. Hou, C. Bian, S. Pan, Y. Li, X. Liu, W. Wang, B. Servin, B. Sayre, B. Zhu, D. Sweney, R. Moore, W. Nie, Y. Shen, R. Zhao, G. Zhang, J. Li, T. Faraut, J. Womack, Y. Zhang, J. Kijas, N. Cockett, X. Xu, S. Zhao, J. Wang, and W. Wang. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nature biotechnology*, 31(2):135–41, February 2013. ISSN 1546-1696. doi: [10.1038/nbt.2478](https://doi.org/10.1038/nbt.2478). URL <http://dx.doi.org/10.1038/nbt.2478>.
- [64] A. Castro-Alonso, C. De la Fe, A. Espinosa de los Monteros, F. Rodríguez, M. Andrada, J. B. Poveda, and P. Herráez. Chronological and immunohistochemical characterization of the mammary immunoinflammatory response in experimental caprine contagious agalactia. *Veterinary Immunology and Immunopathology*, 136(1):43–54, 2009. URL <http://www.sciencedirect.com/science/article/pii/S0165242710000358>.

- [65] J. Ogorevc, S. Prpar, and P. Dovč. In vitro mammary gland model: establishment and characterization of a caprine mammary epithelial cell line. *Acta agriculturae Slovenica*, 94(2):133–138, 2009. ISSN 1176-9343. doi: [10.4137/EBO.S8897](https://doi.org/10.4137/EBO.S8897). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3290115&tool=pmcentrez&rendertype=abstract>.
- [66] J. Ogorevc, S. Prpar, and P. Dovč. Preučevanje molekularnih interakcij med gostiteljem in patogenom pri infekciji epitelnih celic mlečne žleze kože z bakterijo *Mycoplasma agalactiae* (Preliminarni rezultati). In U. Potočnik, editor, *Kolokvij iz genetike*, pages 14–21, Ljubljana, 2010. Slovensko genetsko društvo.
- [67] E. Cifrian, A. J. Guidry, C. N. O'Brien, J. E. Keys, and W. W. Marquardt. Bovine mammary teat and ductal epithelial cell cultures. *American journal of veterinary research*, 55(2):239–46, February 1994. ISSN 0002-9645. URL <http://www.ncbi.nlm.nih.gov/pubmed/8172414>.
- [68] S. H. Kleven and N. Ferguson-Noel. *Mycoplasma synoviae* infection. In Y. M. Saif, A. M. Fadly, J. R. Glisson, L. R. McDougald, N. K. Nolan, and D. E. Swayne, editors, *Diseases of Poultry*, chapter Chapter 21, pages 845–846. Blackwell Publishing, Ames, Iowa, 12th edition, 2008.
- [69] M. Morgan, S. Anders, M. Lawrence, P. Aboyun, H. Pages, and R. Gentleman. [ShortRead]: a [B]ioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*, 25:2607–2608, 2009. doi: [10.1093/bioinformatics/btp450](https://doi.org/10.1093/bioinformatics/btp450). URL <http://dx.doi.org/10.1093/bioinformatics/btp450>.
- [70] C. A. Mathee and S. K. Davis. Molecular insights into the evolution of the family Bovidae: a nuclear DNA perspective. *Molecular biology and evolution*, 18(7):1220–30, July 2001. ISSN 0737-4038. URL <http://www.ncbi.nlm.nih.gov/pubmed/11420362>.
- [71] A. Ropiquet and A. Hassanin. Molecular phylogeny of caprines (Bovidae, Antilopinae): the question of their origin and diversification during the Miocene. *Journal of Zoological Systematics and Evolutionary Research*, 43(1):49–60, February 2005. ISSN 0947-5745. doi: [10.1111/j.1439-0469.2004.00290.x](https://doi.org/10.1111/j.1439-0469.2004.00290.x). URL <http://doi.wiley.com/10.1111/j.1439-0469.2004.00290.x>.
- [72] F. Bibi, M. Bukhsianidze, A. W. Gentry, D. Geraads, and E. S. V. Dimitris S. Kostopoulos. The Fossil Record and Evolution of Bovidae: State of the Field. *Palaentologia Electronica*, 12(3), 2009. URL [http://www.researchgate.net/publication/237365138\\_The\\_Fossil\\_Record\\_and\\_Evolution\\_of\\_Bovidae\\_State\\_of\\_the\\_Field](http://www.researchgate.net/publication/237365138_The_Fossil_Record_and_Evolution_of_Bovidae_State_of_the_Field).
- [73] I. A. Arif, M. A. Bakir, and H. A. Khan. Inferring the phylogeny of bovidae using mitochondrial DNA sequences: resolving power of individual genes relative to complete genomes. *Evolutionary bioinformatics online*, 8:139–50, January 2012. ISSN 1176-9343. doi: [10.4137/EBO.S8897](https://doi.org/10.4137/EBO.S8897). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3290115&tool=pmcentrez&rendertype=abstract>.
- [74] D. Kim, G. Percea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*, 14(4):R36, April 2013. ISSN 1465-6914. doi: [10.1186/gb-2013-14-4-r36](https://doi.org/10.1186/gb-2013-14-4-r36). URL <http://genomebiology.com/2013/14/4/R36>.
- [75] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3):R25, January 2009. ISSN 1465-6914. doi: [10.1186/gb-2009-10-3-r25](https://doi.org/10.1186/gb-2009-10-3-r25). URL <http://genomebiology.com/2009/10/3/R25>.
- [76] B. Langmead and S. L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4):357–9, April 2012. ISSN 1548-7105. doi: [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3322381&tool=pmcentrez&rendertype=abstract>.
- [77] C. Trapnell, B. A. Williams, G. Percea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, and L. Pachter. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology*, 28(5):511–5, May 2010. ISSN 1546-1696. doi: [10.1038/nbt.1621](https://doi.org/10.1038/nbt.1621). URL <http://dx.doi.org/10.1038/nbt.1621>.
- [78] T. M. Lowe and S. R. Eddy. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Research*, 25(5):955–964, March 1997. ISSN 0305-1048. doi: [10.1093/nar/25.5.955](https://doi.org/10.1093/nar/25.5.955). URL <http://nar.oxfordjournals.org/content/25/5/955.abstract?sid=c28dd9c3-943e-44d4-99fe-15ee8a5ac051>.
- [79] A. Conesa, S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics (Oxford, England)*, 21(18):3674–6, September 2005. ISSN 1367-4803. doi: [10.1093/bioinformatics/bti610](https://doi.org/10.1093/bioinformatics/bti610). URL <http://www.ncbi.nlm.nih.gov/pubmed/16081474>.
- [80] S. Anders and W. Huber. Differential expression analysis for sequence count data. *Genome Biology*, 11:R106, 2010. doi: [10.1186/gb-2010-11-10-r106](https://doi.org/10.1186/gb-2010-11-10-r106). URL <http://genomebiology.com/2010/11/10/R106/>.
- [81] M. D. Robinson, D. J. McCarthy, and G. K. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1):139–40, January

2010. ISSN 1367-4811. doi: [10.1093/bioinformatics/btp616](https://doi.org/10.1093/bioinformatics/btp616). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2796818&tool=pmcentrez&rendertype=abstract>.
- [82] K. A. Baggerly, L. Deng, J. S. Morris, and C. M. Aldaz. Differential expression in SAGE: accounting for normal between-library variation. *Bioinformatics (Oxford, England)*, 19(12):1477–83, August 2003. ISSN 1367-4803. URL <http://www.ncbi.nlm.nih.gov/pubmed/12912827>.
- [83] K. K. Abu-Amero, E. A. Abu-Groun, M. A. Halablab, and R. J. Miles. Kinetics and distribution of alcohol oxidising activity in *Acholeplasma* and *Mycoplasma* species. *FEMS microbiology letters*, 183(1): 147–51, February 2000. ISSN 0378-1097. URL <http://www.ncbi.nlm.nih.gov/pubmed/10650218>.
- [84] A. J. Westermann, S. A. Gorski, and J. Vogel. Dual RNA-seq of pathogen and host. *Nature reviews. Microbiology*, 10(9):618–30, September 2012. ISSN 1740-1534. doi: [10.1038/nrmicro2852](https://doi.org/10.1038/nrmicro2852). URL <http://dx.doi.org/10.1038/nrmicro2852>.
- [85] J. Ernst and Z. Bar-Joseph. STEM: a tool for the analysis of short time series gene expression data. *BMC bioinformatics*, 7:191, January 2006. ISSN 1471-2105. doi: [10.1186/1471-2105-7-191](https://doi.org/10.1186/1471-2105-7-191). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1456994&tool=pmcentrez&rendertype=abstract>.
- [86] K. M. McDermott, B. Y. Liu, T. D. Tlsty, and G. J. Pazour. Primary cilia regulate branching morphogenesis during mammary gland development. *Current biology : CB*, 20(8):731–7, April 2010. ISSN 1879-0445. doi: [10.1016/j.cub.2010.02.048](https://doi.org/10.1016/j.cub.2010.02.048). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2916967&tool=pmcentrez&rendertype=abstract>.
- [87] M. D. Young, M. J. Wakefield, G. K. Smyth, and A. Oshlack. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology*, 11(2):R14, January 2010. ISSN 1465-6914. doi: [10.1186/gb-2010-11-2-r14](https://doi.org/10.1186/gb-2010-11-2-r14). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2872874&tool=pmcentrez&rendertype=abstract>.
- [88] S. Tsai. Mycoplasmas and Oncogenesis: Persistent Infection and Multistage Malignant Transformation. *Proc Natl Acad Sci U S A*, 92(22): 10197–10201, October 1995. ISSN 0027-8424. doi: [10.1073/pnas.92.22.10197](https://doi.org/10.1073/pnas.92.22.10197). URL <http://www.pnas.org/content/92/22/10197.short>.
- [89] S. Feng, S. Tsai, J. Rodriguez, and S. Lo. Mycoplasma Infections Prevent Apoptosis and Induce Malignant Transformation of Interleukin-3-Dependent 32D Hematopoietic Cells. *Mol. Cell. Biol.*, 19(12):7995–8002, December 1999. URL <http://mcb.asm.org/content/19/12/7995.short>.
- [90] K. Namiki, S. Goodison, S. Porvasnik, R. W. Allan, K. A. Izkowski, C. Urbanek, L. Reyes, N. Sakamoto, and C. J. Rosser. Persistent exposure to *Mycoplasma* induces malignant transformation of human prostate cells. *PLoS One*, 4(9):e6872, January 2009. ISSN 1932-6203. doi: [10.1371/journal.pone.0006872](https://doi.org/10.1371/journal.pone.0006872). URL <http://dx.plos.org/10.1371/journal.pone.0006872>.
- [91] A. H. Chang and J. Parsonnet. Role of bacteria in oncogenesis. *Clinical microbiology reviews*, 23(4):837–57, October 2010. ISSN 1098-6618. doi: [10.1128/CMR.00012-10](https://doi.org/10.1128/CMR.00012-10). URL <http://cmr.asm.org/content/23/4/837.short>.
- [92] M. B. Rogers. Mycoplasma and cancer: in search of the link. *Oncotarget*, 2(4):271–3, April 2011. ISSN 1949-2553. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3248167&tool=pmcentrez&rendertype=abstract>.
- [93] G. L. Nicolson, M. Y. Nasralla, J. Haier, and J. Pomfret. High frequency of systemic mycoplasmal infections in Gulf War veterans and civilians with Amyotrophic Lateral Sclerosis (ALS). *Journal of Clinical Neuroscience*, 9(5):525–529, 2002. URL <http://www.sciencedirect.com/science/article/pii/S0967586801910756>.
- [94] M. Pehlivan, G. Itirli, H. Onay, H. Bulut, M. Koyuncuoglu, and S. Pehlivan. Does *Mycoplasma* sp. play role in small cell lung cancer? *Lung cancer (Amsterdam, Netherlands)*, 45(1):129–30, July 2004. ISSN 0169-5002. doi: [10.1016/j.lungcan.2004.01.007](https://doi.org/10.1016/j.lungcan.2004.01.007). URL <http://www.ncbi.nlm.nih.gov/pubmed/15196745>.
- [95] Stéphane Belin, Anne Beghin, Eduardo Solano-González, Laurent Bezin, Stéphanie Brunet-Manquat, Julien Textoris, Anne-Catherine Prats, Hichem C Mertani, Charles Dumontet, and Jean-Jacques Diaz. Dysregulation of ribosome biogenesis and translational capacity is associated with tumor progression of human breast cancer cells. *PLoS One*, 4(9):e7147, January 2009. ISSN 1932-6203. doi: [10.1371/journal.pone.0007147](https://doi.org/10.1371/journal.pone.0007147). URL <http://dx.plos.org/10.1371/journal.pone.0007147>.
- [96] S. Kleinschmidt, J. Spersger, R. Rosengarten, and M. Hewicker-Trautwein. Long-term survival of *Mycoplasma bovis* in necrotic lesions and in phagocytic cells as demonstrated by transmission and immunogold electron microscopy in lung tissue from experimentally infected calves. *Veterinary microbiology*, 162(2-4):949–53, March 2013. ISSN 1873-2542. doi: [10.1016/j.vetmic.2012.11.039](https://doi.org/10.1016/j.vetmic.2012.11.039). URL <http://www.ncbi.nlm.nih.gov/pubmed/23294861>.

- [97] K. M. Swanson, K. Stelwagen, J. Dobson, H. V. Henderson, S. R. Davis, V.C. Farr, and K. Singh. Transcriptome profiling of *Streptococcus uberis*-induced mastitis reveals fundamental differences between immune gene expression in the mammary gland and in a primary cell culture model. *Journal of Dairy Science*, 92(1):117–129, 2009. URL <http://www.sciencedirect.com/science/article/pii/S0022030209703169>.