UNIVERZA V LJUBLJANI BIOTEHNIŠKA FAKULTETA

Janez KOKOŠAR

## PRIMERJALNA GENOMIKA IN ANALIZA VLOG GENOV SESALCEV, NASTALIH IZ RETROELEMENTOV

DOKTORSKA DISERTACIJA

Ljubljana, 2013

UNIVERZA V LJUBLJANI BIOTEHNIŠKA FAKULTETA

Janez KOKOŠAR

## PRIMERJALNA GENOMIKA IN ANALIZA VLOG GENOV SESALCEV, NASTALIH IZ RETROELEMENTOV

DOKTORSKA DISERTACIJA

## COMPARATIVE GENOMICS AND PHENOTYPIC IMPACT OF NOVEL RETROTRANSPOSON-DERIVED GENES IN PLACENTAL MAMMALS

DOCTORAL DISSERTATION

Ljubljana, 2013

Doktorsko delo je bilo opravljeno na Odseku za molekularne in biomedicinske znanosti Instituta »Jožef Stefan« v Ljubljani.

Na podlagi Statuta Univerze v Ljubljani ter po sklepu Senata Biotehniške fakultete in sklepa Senata Univerze z dne 7.12.2011 je bilo potrjeno, da kandidat izpolnjuje pogoje za neposredni prehod na doktorski Podiplomski študij bioloških in biotehniških znanosti ter opravljanje doktorata znanosti s področja genetike. Za mentorja je bil imenovan doc. dr. Dušan Kordiš.

Mentor: doc. dr. Dušan KORDIŠ

Komisija za oceno iz zagovor:

- Predsednica: prof. dr. Branka JAVORNIK Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za agronomijo
- Član: prof. dr. Simon HORVAT Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko
- Član: doc. dr. Dušan KORDIŠ Institut »Jožef Stefan«, Odsek za molekularne in biomedicinske znanosti

Datum zagovora: 17. 7. 2013

Naloga je rezultat lastnega raziskovalnega dela. Podpisani se strinjam z objavo svoje naloge v polnem tekstu na spletni strani Digitalne knjižnice Biotehniške fakultete. Izjavljam, da je naloga, ki sem jo oddal v elektronski obliki, identična tiskani verziji.

Doktorand: Janez KOKOŠAR

#### KLJUČNA DOKUMENTACIJSKA INFORMACIJA

#### ŠD Dd

- DK UDK 575: 577.21 (043.3) = 163.6
- KG filogenomska analiza/udomačeni geni/retrotranspozoni/molekularna udomačitev/placentalni sesalci/neofunkcionalizacija
- AV KOKOŠAR, Janez, univ. dipl. biokemik
- SA KORDIŠ, Dušan (mentor)
- KZ SI-1000 Ljubljana, Jamnikarjeva 101
- ZA Univerza v Ljubljani, Biotehniška fakulteta, Podiplomski študij bioloških in biotehniških znanosti, znanstveno področje genetika
- LI 2013
- IN PRIMERJALNA GENOMIKA IN ANALIZA VLOG GENOV SESALCEV, NASTALIH IZ RETROELEMENTOV
- TD doktorska disertacija
- OP XII, 112 str., 12 pregl., 20 sl., 3 pril., 179 vir.
- IJ sl
- JI sl/en
- Genomi vretenčarjev, predvsem sesalcev, vsebujejo številne gene, ki so v procesu AI molekularne udomačitve (domestikacije) nastali iz različnih transpozicijskih elementov. Zaradi omejenega taksonomskega vzorčenja do sedaj še nista bila zadovoljivo pojasnjena nastanek in evolucija teh genov. Za določitev izvora in evolucije multigenskih družin udomačenih genov, nastalih iz retroelementov družine Ty3/Gypsy (Metaviridae) smo uporabili različne metode filogenetske analize in široko taksonomsko vzorčenje (preiskali smo več kot 90 različnih genomov strunarjev). Dokazali smo, da je večina udomačenih genov nastala v predniku Eutheria, pred ločitvijo štirih placentalnih nadredov. S filogenomsko analizo udomačenih genov in njihovih izvornih retroelementov smo dokazali, da so udomačeni geni nastali kot posledica več neodvisnih dogodkov molekularne udomačitve, nastale genske družine pa so se povečevale z genskimi duplikacijami. Dokazali smo, da so tako udomačeni geni kot njihove kromosomske lokacije močno ohranjeni znotraj placentalnih sesalcev, kar nakazuje na močno adaptivno evolucijo in hitro fiksacijo teh genov v predniku Eutheria. Dokazali smo, da so udomačeni geni nastali iz ostankov nekoč aktivnih transpozicijskih elementov, ter da so svoje regulatorne in promotorske regije pridobili na novo. Identifikacija vezavnih partnerjev človeškega proteina PNMA5 nam je podala nov vpogled v pridobivanje bioloških vlog udomačenih genov, katerih nastanek je bil najverjetneje povezan z razvojem pomembnih evolucijskih novosti pri placentalnih sesalcih.

#### **KEY WORDS DOCUMENTATION**

#### ND Dd

- DC UDC 575: 577.21 (043.3) = 163.6
- CX phylogenomic analysis/domesticated genes/retrotransposons/molecular domestication/placental mammals/neofunctionalization
- AU KOKOŠAR, Janez
- AA KORDIŠ, Dušan (supervisor)
- PP SI-1000 Ljubljana, Jamnikarjeva 101
- PB University of Ljubljana, Biotechnical Faculty, Postgraduate Study of Biological and Biotechnical Sciences, Field: Genetics
- PY 2013
- TI COMPARATIVE GENOMICS AND PHENOTYPIC IMPACT OF NOVEL RETROTRANSPOSON-DERIVED GENES IN PLACENTAL MAMMALS
- DT doctoral dissertation
- NO XII, 112 p., 12 tab., 20 fig., 3 ann., 179 ref.
- LA sl
- AL sl/en
- Vertebrates, especially mammals, possess numerous single copy genes that AB originated from diverse multicopy retroelements in the process of molecular domestication. Previous studies provided limited insight into evolutionary relationships between different families of domesticated genes (DGs), either due to insufficient taxonomic sampling or the analysis of a single family of domesticated genes only. We traced the genesis, evolution and regulatory wiring of the retroelement-derived domesticated genes (RDDGs) through phylogenomic analysis, using whole-genome information from more than 90 chordate genomes. Phylogenomic analysis of the DGs in chordate genomes provided direct evidence that major diversification has occurred only in the ancestor of placental mammals. Mammalian RDDGs have been shown to originate in several steps by independent domestication events and to diversify later by gene duplications. We have shown that domesticated genes and their chromosomal positions were fully established in the ancestor of eutherian mammals. We demonstrated that strong adaptive evolution has diversified domesticated genes before their fixation in the ancestor of placental mammals. By analysis of active Metaviridae lineages in amniotes, we have demonstrated that Eutheria-specific domesticated genes originated from retroelement remains, and that domesticated genes gained their promoter and regulatory regions de novo. Functional characterization of PNMA5 gene provided further insight into the role of domesticated genes in the origin of Eutheria-specific innovations and adaptations.

#### KAZALO VSEBINE

KLJUČNA DOKUMENTACIJSKA INFORMACIJAIII
KEY WORDS DOCUMENTATIONIV
KAZALO VSEBINEV
KAZALO PREGLEDNICIV
KAZALO SLIK X
KAZALO PRILOG XI
OKRAJSAVE IN SIMBOLI
1 UVOD1
1.1 NAMEN IN HIPOTEZE
2 PREGLED OBJAV
2.1 RAZNOLIKOST TRANSPOZICIJSKIH ELEMENTOV 3
2.1.1 Elementi Metaviridae pri vretenčarjih
2.2 VPLIV TRANSPOZICIJSKIH ELEMENTOV NA EVOLUCIJO GENOMOV 5
2.3 KAKO NASTAJAJO NOVI GENI?
2.4 TRANSPOZICIJSKI ELEMENTI KOT VIR NOVIH KODIRAJOČIH
ZAPOREDIJ
2.4.1 Eksonizacija
2.4.2 Molekularna udomačitev (domestikacija)10
2.5 NOVI GENI IN PRIDOBIVANJE PROMOTORSKIH ZAPOREDIJ – PRIMER
2.6 UDOMAČENI GENI NASTALI IZ TRANSPOZICIJSKIH ELEMENTOV 11
2.6.1 Udomačeni geni nastali iz DNA transpozonov
2.6.2 Udomačeni geni nastali iz LTR retroelementov
2.6.2.1 Udomaceni geni nastali iz plascnega proteina (Env)
2.6.2.2 Udomaceni geni nastali iz domene gag
2.6.2.3 Udomaćeni geni nastali iz proteazne domene
2.6.2.4 Udomačeni geni nastali iz integrazne domene
2.7 BIOLOŠKA VLOGA UDOMAČENIH GENOV – RAZVOJ FENOTIPSKIH NOVOSTI

2.8	VPLIV UDOMAČENIH GENOV NA GENSKA REGULATORNA OMREŽJA	A. 19
2.9	PROTEINSKE INTERAKCIJE GAG IN PROTEINOV, PODOBNIH GAG	20
2.10	EVOLUCIJSKI ODNOSI GLAVNIH SKUPIN VRETENČARJEV IN	
SESA	LCEV	20
3 N	IATERIAL IN METODE	22
3.1	MATERIAL	22
3.1.	1 Seznam kemikalij	22
3.1.	2 Uporabljeni encimi, protitelesa in proteini	23
3.1.	3 Seznam uporabljenih gojišč (Y2H)	23
3.1.	4 Uporabljeni sevi kvasovk S. cerevisiae	24
3.1.	5 Uporabljeni bakterijski sevi	25
3.1.	.6 Uporabljene celične linije	25
3.1.	7 Uporabljeni plazmidi	25
3.1.	8 Uporabljeni začetni oligonukleotidi	25
3.2	METODE	26
3.2.	1 Pridobivanje podatkov	26
3.2.	2 Filogenomska analiza udomačenih genov	26
3.2.	<b>3</b> Analiza ohranjenih kromosomskih položajev (analiza sintenije)	26
3.2.	4 Filogenetska analiza udomačenih genov	27
3.2.	5 Filogenetska analiza elementov Ty3/Gypsy pri Deuterostomia	27
3.2.	6 Analiza promotorskih in drugih regulatornih zaporedij	27
3.2.	7 Metode molekulskega kloniranja	28
3.2.	8 Dvohibridni sistem kvasovke (Y2H)	29
3	.2.8.1 Kloniranje PNMA5	29
3	.2.8.2 Testiranje toksičnosti zapisa PNMA5-DB v celicah S. cerevisiae	30
3	.2.8.3 Y2H	30
3.2.	9 Ko-imunoprecipitacija	30
3.2.	.10 Določanje celične lokalizacije proteina PNMA5	31
3.2.	.11 Ko-lokalizacija proteina PNMA5 s PML jedrnimi telesci	31
3.2.	.12 Določanje celične lokalizacije proteina SPTBN1	32
3.2.	13 Luciferazni test	32
4 R	REZULTATI	33
11	ΙΡΕΝΤΙΕΙΚ Α CH Α ΤΙΡΟΜΑ ČΕΝΙΗ CENOV ΝΑ STATΗ 17	
4.1 RETR	$\frac{1}{2} \frac{1}{2} \frac{1}$	33
		55
4.2	TAKSONOMSKA RAZŠIRJENOST UDOMAČENIH GENOV	33
4.0		~ 4
4.3	FILOGENE ISKA ANALIZA UDOMACENIH GENOV	34
4.3.	1 Druzina Susni	35

4. 4. 4.	3.2 3.3 3.4 3.5	Družina PNMA	39 16 17 18
4.4	K	DAJ IN KJE SO NASTALI UDOMAČENI GENI?	51
4.5 GEN	NOV NOV	IUKLEOTIDNA IN AMINOKISLINSKA ZAPOREDJA UDOMAČENIH 7 SO IZJEMNO OHRANJENA5	51
4.6 GEN	A JOV	NALIZA OHRANJENOSTI KROMOSOMSKIH POLOŽAJEV UDOMAČENII 7	H 52
4.7	K	ROMOSOMSKA MOBILNOST UDOMAČENIH GENOV	54
4.8 <i>TY3/</i>	A /GY	NALIZA PROGENITORJEV UDOMAČENIH GENOV: ELEMENTI PSY (METAVIRIDAE) PRI DEUTEROSTOMIA5	57
4.9 RET	S RO	TRUKTURNA ORGANIZACIJA UDOMAČENIH GENOV, NASTALIH IZ ELEMENTOV5	59
4.10 REC <b>4.</b>	E Gul 10.1	DE NOVO PRIDOBIVANJE IN EVOLUCIJA PROMOTORSKIH IN DRUGIH ATORNIH REGIJ UDOMAČENIH GENOV	53 5 <b>3</b>
4.	10.2	2 De novo pridobivanje in evolucija 5' UTR / promotorske regije	56
4.	10.3	b De novo pridobivanje in evolucija 3' UTR regije6	<b>58</b>
4.11 Z D	E VOI	OLOČANJE VEZAVNIH PARTNERJEV ČLOVEŠKEGA PROTEINA PNMA HIBRIDNIM SISTEMOM KVASOVKE (Y2H)6	5 59
4.	11.1 11 7	. Test odvisnosti aktivacije reporterskih genov od prisotnosti PNMA5	/1
in	11.2 1un	oprecipitacijo	12
4.12	S	UB-CELIČNA LOKALIZACIJA PROTEINA PNMA57	73
4.13	S	UB-CELIČNA LOKALIZACIJA PROTEINA SPTBN1	15
4.14 PRC 5	A DTE RA	KTIVNOST SMAD SIGNALNE POTI V ODVISNOSTI OD IZRAŽANJA INA PNMA5	15 17
5.1	A	NALIZA RAZŠIRJENOSTI UDOMAČENIH GENOV	17
5.2	F	ILOGENOMSKA ANALIZA UDOMAČENIH GENOV7	78
5.3 UDC	LZ DMA	ZJEMNA OHRANJENOST ZAPOREDIJ IN KROMOSOMSKIH POLOŽAJEV AČENIH GENOV	30

8	VIRI	96
7.2	SUMMARY	93
7.1	POVZETEK	92
7	POVZETEK (SUMMARY)	92
5.8 6	BIOLOŠKA VLOGA PROTEINA PNMA5 SKLEPI	85 <b>89</b>
5.7	PRIDOBIVANJE NOVIH FENOTIPSKIH VLOG UDOMAČENIH GENOV	84
5.6 ZAl	DE NOVO PRIDOBIVANJE PROMOTORSKIH IN DRUGIH REGULATORN POREDIJ	√IH 83
5.5 (EL	FILOGENETSKA ANALIZA PROGENITORJEV UDOMAČENIH GENOV EMENTOV METAVIRIDAE)	81
5.4	KROMOSOMSKA MOBILNOST UDOMAČENIH GENOV	81

ZAHVALA PRILOGE

#### **KAZALO PREGLEDNIC**

str
<b>Pregl. 1:</b> Poznane biološke vloge genov/proteinov iz družine Sushi
<b>Pregl. 2:</b> Poznane biološke vloge genov/proteinov iz družine PNMA
Pregl. 3: Seznam dodatkov minimalnim gojiščem
Pregl. 4: Uporabljeni sevi S. cerevisiae
Pregl. 5: Uporabljene celične linije
Pregl. 6: Prisotnost udomačenih genov znotraj posameznih taksonomskih skupin
Pregl. 7: Primerjava AK ohranjenosti med proteini pri človeku (Homo sapiens) in
ortolognimi proteini pri vretenčarjih za nekatere udomačene gene
Pregl. 8: Razširjenost različnih skupin Metaviridae pri tetrapodih
Pregl. 9: Struktura udomačenih genov
Pregl. 10: Analiza promotorskih regij udomačenih genov
Pregl. 11: Dolžina 3'-UTR regij (bp) pri človeških in mišjih udomačenih genih v povezavi
z njihovim profilom izražanja69
Pregl. 12: Identificirani potencialni vezavni partnerji človeškega proteina PNMA570

#### KAZALO SLIK

Sl. 1: Bayesova analiza evolucijskih odnosov med geni iz družine Sushi	
Sl. 2: NJ-analiza evolucijskih odnosov med geni iz družine Sushi	
Sl. 3: Bayesova analiza evolucijskih odnosov med geni iz družine PNMA	41
Sl. 4: ML-analiza evolucijskih odnosov med geni iz družine PNMA	43
Sl. 5: NJ-analiza evolucijskih odnosov med geni iz družine PNMA	45
Sl. 6: NJ-analiza evolucijskih odnosov med geni ARC	47
Sl. 7: NJ-analiza evolucijskih odnosov med geni ASPRV1	
Sl. 8: NJ-analiza evolucijskih odnosov v integrazni družini udomačenih genov	50
Sl. 9: Shematski prikaz ohranjenosti sintenije genov družine Sushi	53
Sl. 10: Kromosomska mobilnost udomačenih genov	56
Sl. 11: NJ analiza elementov Metaviridae pri Deuterostomia	58
Sl. 12: Struktura udomačenih genov nastalih iz retroelementov	60
Sl. 13: Ključni mehanizmi vključeni v neofunkcionalizacijo novonastalih udomačer	nih
genov	65
Sl. 14: Različni tipi promotorjev pri udomačenih genih	66
Sl. 15: Shematski prikaz strukture proteina SPTBN1	71
Sl. 16: Test odvisnosti aktivacije reporterskih genov od prisotnosti PNMA5	72
Sl. 17: Sub-celična lokalizacija človeškega proteina PNMA5-EGFP v celicah HEKZ	293T 73
Sl. 18: Protein PNMA5-EGFP je ko-lokaliziran z jedrnimi telesci PML	74
Sl. 19: Sub-celična lokalizacija proteina SPTBN1	75
Sl. 20: Aktivnost SMAD signalne poti v odvisnosti od izražanja proteina PNMA5 v	celicah
HEK293T	

#### **KAZALO PRILOG**

- **PRILOGA A**: Seznam genov uporabljenih pri filogenomskih analizah
- **PRILOGA B**: Shematski prikaz ohranjenosti sintenije udomačenih genov
- **PRILOGA C:** Vezavna mesta transkripcijskih faktorjev v promotorjih udomačenih genov pri človeku in miši.

## OKRAJŠAVE IN SIMBOLI

AGE	agarozna gelska elektroforeza
AP	aspartatna proteinaza
BSA	albumin govejega seruma
BP	bazni par
DB	baza podatkov (ang. <i>database</i> )
cDNA	komplementarna DNA
ENV	plaščni protein Env
EST	oznaka izraženega zaporedja (ang. expressed sequence tag)
FBS	goveji serum
GAG	(proteinska) domena Gag
GSS	ang. genome survey sequences
HTGS	ang. high throughput genome sequences
IN	integraza
LB	gojitveni medij po Luria-Bertaniju
LINE	dolgi razpršeni jedrni elementi
LTR	dolga končna ponovitev (ang. long terminal repeat)
ML	metoda največjega verjetja (ang. maximum likelihood)
NCBI	»National Center for Biotechnology Information«
NJ	metoda združevanja sosedov (ang. neighbor joining)
ORF	odprt bralni okvir
PBS	fosfatni pufer
PCR	verižna reakcija s polimerazo
PPT	polipurinski trakt
RH	RNA-za H
RT	reverzna transkriptaza
TE	transpozicijski element
TSD	podvojitev tarčnega mesta (ang. target site duplication)
UTR	neprevedena regija (ang. untranslated region)
WGS	zaporedja celotnih genomov
Y2H	kvasni dvo-hibridni sistem (ang. yeast two hybrid system)
X-α-Gal	5-bromo-4-kloro-3-indoil-α-D-galaktopiranozid

## 1 UVOD

Genomi vretenčarjev, predvsem pa sesalcev, vsebujejo mnoga kodirajoča zaporedja, ki so nastala iz transpozicijskih elementov (Zdobnov in sod., 2005; Campillos in sod., 2006, Volff, 2006). Zaporedja retrotranspozonov, ki kodirajo različne zapise gag, pol in env, so med molekularno udomačitvijo (domestikacijo) pogosto pridobila sposobnost kodiranja novih funkcionalnih proteinov (Miller in sod., 1999; Zdobnov in sod., 2005; Volff, 2006). Udomačeni geni, ki so nastali iz transpozicijskih elementov, predstavljajo novo in stabilno komponento genomov, ki pogosto opravlja pomembno vlogo (Miller in sod., 1999). Nastanek in evolucija novih udomačenih genov, nastalih iz družine retroelementov Ty3/Gypsy (Metaviridae), sta bila do sedaj le delno pojasnjena. Pomanjkljivost dosedanjih študij je bilo predvsem omejeno taksonomsko vzorčenje (Lloréns in Marín, 2001; Zdobnov in sod., 2005; Campillos in sod., 2006), oziroma je bilo zadovoljivo vzorčenje uporabljeno za eno samo družino udomačenih genov (Lynch in sod., 2003; Brandt in sod., 2005a; Brandt in sod., 2005b). Dosedanje študije zato niso podale celovitega vpogleda v nastanek, razširjenost in evolucijo genov, ki so nastali iz retroelementov.

#### 1.1 NAMEN IN HIPOTEZE

Obstoječe študije niso podale celovitega odgovora glede razširjenosti, evolucije in funkcionalne diverzifikacije udomačenih genov, nastalih iz retroelementov. Ravno tako so ostala nepojasnjena vprašanja glede njihovega izvora, časa in mehanizma nastanka (molekularne udomačitve). Osnovni cilji doktorske disertacije so tako pojasnitev izvora, evolucije in mehanizma nastanka multigenskih družin udomačenih genov. Določili bomo razširjenost udomačenih genov in čas njihovega nastanka. Pojasnili bomo evolucijske odnose znotraj različnih multigenskih družin udomačenih genov. Določitev razširjenosti in evolucijske dinamike njihovih izvornih elementov (retrotranspozonov) je ključnega pomena za razumevanje mehanizma nastanka udomačenih genov ter mehanizma pridobivanja njihovih regulatornih in promotorskih zaporedij. Dobro pojasnjeni evolucijski odnosi znotraj različnih družin udomačenih genov so ključni tudi za določitev dinamike njihove kromosomske mobilnosti. Pojasnitev evolucije, mehanizma nastanka in pridobivanja regulatornih zaporedij teh genov je pomembna za razumevanje njihove funkcionalne diverzifikacije in pridobivanja novih bioloških vlog.

Predvidevamo, da bomo z analizo vseh dostopnih sesalskih in drugih vretenčarskih genomov pridobili poglobljen vpogled v razširjenost in evolucijo različnih multigenskih družin udomačenih genov, ter razširjenost njihovih izvornih transpozicijskih elementov. Iskanje zaporedij udomačenih genov in njihovih izvornih elementov bomo izvajali znotraj različnih bioloških podatkovnih zbirk. Iskanje bomo izvajali iterativno z uporabo orodja BLAST. Pri iskanju bomo uporabili različna zaporedja anotiranih udomačenih genov, ter različna zaporedja njihovih izvornih

elementov. Evolucijske odnose znotraj multigenskih družin udomačenih genov bomo določili z uporabo različnih metod filogenetske analize.

Analiza genomov sesalcev iz vseh štirih placentalnih nadredov (Laurasiatheria, Euarchontoglires, Afrotheria, Xenarthra) in iz skupin sesalcev Prototheria in Metatheria nam bo omogočila odgovoriti na vprašanje o izvoru sesalskih udomačenih genov ter o času njihove molekularne udomačitve. Uporaba širokega taksonomskega vzorčenja je ključna za odgovore na vprašanja glede izvora, razširjenosti in časa nastanka udomačenih genov prisotnih pri sesalcih. Primerjava razširjenosti udomačenih genov med različnimi skupinami placentalnih sesalcev, prednikom placentalnih sesalcev in pri bazalnih sesalcih je ključna za razumevanje s pojavom udomačenih genov povezanega razvoja fenotipskih novosti tekom evolucije placentalnih sesalcev.

Predvidevamo, da bomo z analizo velike količine podatkov pridobili informacije tako o strukturni organizaciji udomačenih genov pri sesalcih, kot o sestavi njihovih regulatornih regij. To nam bo omogočilo pojasniti nastanek, regulatorno evolucijo in funkcionalno diverzifikacijo teh genov. Za strukturno in funkcionalno analizo udomačenih genov bomo uporabili različne pristope filogenomske in bioinformatske analize. Strukturno organizacijo udomačenih genov in njihovih regulatornih regij bomo analizirali na osnovi podatkov v javnih podatkovnih bazah. Analizirali bomo velikost genov, velikost 3'- in 5'-neprevedenih zaporedij in prisotnost intronov znotraj kodirajočih ali nekodirajočih zaporedij. S pomočjo orodij za prepoznavanje sekvenčnih motivov bomo prepoznali strukturne značilnosti promotorskih zaporedij in prisotnost vezavnih mest za transkripcijske faktorje. Analizirali bomo strukturno organizacijo (prisotnost proteinskih domen) genskih produktov udomačenih genov. Analizirali bomo lokacijo in kromosomsko organizacijo udomačenih genov. Iz podatkovnih zbirk bomo pridobili podatke o izraženosti udomačenih genov (EST, mikromreže) pri različnih vrstah in v različnih tkivih.

Predvidevamo, da bomo z uporabo dvo-hibridnega sistema kvasovke (Y2H) pridobiti podatke o proteinskih interaktorjih človeškega proteina PNMA5. Analiza teh podatkov bo podala nov vpogled v vlogo in funkcijo proteina PNMA5 pri placentalnih sesalcih. Za izvedbo presejalnega testa Y2H bomo uporabili cDNA knjižnico pred pripravljeno na osnovi možganskega tkiva človeka. Rezultate presejalnega testa bomo validirali z uporabo ko-imunoprecipitacije ter z uporabo konfokalne mikroskopije (ko-lokalizacija).

## 2 PREGLED OBJAV

Transpozicijski elementi (TE) so mobilni genetski elementi, ki imajo sposobnost premikanja in replikacije znotraj genoma gostitelja, v primeru horizontalnega prenosa pa tudi med različnimi genomi (Kidwell in Lisch, 2000; Kidwell, 2001; Kazazian, 2004; Goodier, 2008). Z dostopnostjo velike količine genomskih podatkov se je poznavanje razširjenosti in zastopanosti različnih skupin TE, predvsem pri različnih skupinah evkariontov, bistveno izboljšalo. Delež TE in ostankov nekoč aktivnih TE je lahko v genomih gostiteljev zelo velik - pogosto več kot 50 % genoma organizma (Wicker in sod., 2007). Še posebno visok je v genomih nekaterih rastlin in živali (npr. koruza, 60 %; človek, 45 %), v manjših genomih (npr. *S. cerevisiae*) pa je delež TE bistveno manjši (3-5 %) (Biémont in Vieira, 2006). TE so prisotni tako pri evkariontih, kjer predstavljajo prevladujočo komponento genoma, kot prokariontih. Aktivnost in evolucijska dinamika TE sta pomembna dejavnika, ki vplivata na velikost gostiteljskega genoma, saj je povečano število kopij TE pogosto sorazmerno s povečano velikostjo genoma (Kidwell, 2002).

TE so bili sprva prepoznani kot sebični, škodljivi in odvečni genomski elementi, ki lahko s svojo aktivnostjo in mobilnostjo povzročajo škodljive učinke v genomu gostitelja, negativno vplivajo na fitnes gostiteljskega organizma in s tem bistveno zmanjšajo njegovo sposobnost preživetja (Doolittle in Sapienza, 1980). Aktivnost TE se v stresnih pogojih poveča, insercija/translokacija TE v aktivne genomske lokuse lahko povzroči spremembe v izražanju in aktivnosti genov (Capy in sod., 2000), sproži pa lahko tudi velike genomske preureditve, kot so podvojitve, inverzije in translokacije. Čeprav ima aktivnost TE pogosto negativen in škodljiv vpliv na posamezen organizem oziroma celico, pa je njihova evolucijska vloga bistveno pomembnejša (Kidwell in Lisch, 2000; Kidwell, 2001; Kazazian, 2004; Böhne in sod., 2008; Oliver in Greene, 2009). Aktivnost TE predstavlja enega od najpomembnejših dejavnikov, ki vplivajo na spremembe v organizaciji in na evolucijo genomov. S svojim mutagenim delovanjem TE bistveno prispevajo k variabilnosti na ravni genov, organizacije kromosomov oziroma celotnih genomov, in genskih regulatornih omrežij (Feschotte, 2008). TE s svojim delovanjem vplivajo na številne celične procese ter so pomembno sodelovali pri evoluciji in preoblikovanju vretenčarskih genomov, hkrati pa so vedno bolj prepoznani kot pomemben vir novih kodirajočih zaporedij v obliki eksonov in celotnih genov, s katerimi so obogatili gostiteljev genski nabor (Volff, 2006; Böhne in sod., 2008, Kaessmann, 2010).

#### 2.1 RAZNOLIKOST TRANSPOZICIJSKIH ELEMENTOV

Transpozicijski elementi so glede na uporabljeno strategijo za replikacijo in premikanje v genomu uvrščeni v dve skupini. Elementi skupine 1 oziroma DNA-transpozoni se po genomu gostitelja premikajo v obliki zaporedij DNA, njihovo mobilnost pa uravnava

encim transpozaza. Transpozaza je kodirana znotraj odprtega bralnega okvirja na avtonomnih DNA-transpozonih, na obeh koncih DNA-transpozonov pa so obrnjene ponovitve (TIR – ang. »terminal inverted repeat«). Encim transpozaza mobilni element izreže iz izvornega mesta, element pa se nato v tarčno mesto integrira brez predhodne podvojitve (mehanizem cut & paste) (Kazazian, 2004). Med avtonomne DNA-transpozicijske elemente spadajo elementi naddružin Tc1/mariner, Zator, Merlin, PIF/Harbinger, MULE, P, hAT, Kolobok, Novosib, PiggyBac, Sola1/2/3, CMC, Transib, Academ in Ginger (Yuan in Wessler, 2011), med neavtonomne pa elementi iz družin Helitron in Polintron (Maverick) (Wicker in sod., 2007; Kapitonov in Jurka, 2008).

TE skupine 2 oziroma retrotranspozoni se po genomu širijo preko RNA intermediatov. Prepišejo se v molekulo RNA, nato pa se s pomočjo reverzne transkriptaze (RT) prevedejo nazaj v molekulo DNA in se znova vstavijo v genom. Vsak zaključen cikel retrotranspozicije tako proizvede novo kopijo retroelementa. Posledično retrotranspozoni pogosto predstavljajo večino ponavljajočih se zaporedij v velikih genomih. Elementi skupine 2 so razdeljeni na elemente z dolgimi končnimi ponovitvami (LTR-retrotranspozoni) in na elemente brez LTR-zaporedij (ne-LTR retrotranspozoni) (Kazazian, 2004).

LTR retrotranspozononi so ponavadi sestavljeni iz zapisov gag, pol ter v nekaterih primerih tudi env, ki jih na 5'- in 3'-koncih obdaja LTR-zaporedje (Kazazian, 2004). LTR-zaporedja so potrebna za pravilno integracijo elementa v genom, vsebujejo pa tudi različna promotorska in ojačevalna zaporedja, ki vplivajo na izražanje proteinov, kodiranih na retrotranspozonu. Tipični strukturni elementi LTR-retrotranspozonov so: a) 4-6 bp dolge podvojitve tarčnega mesta (TSD – ang. »target site duplication«) na 5'- in 3'- koncih; b) PBS (ang. »primer bindig site«), ki se nahaja za 5'- LTR; c) polipurinski trakt (PPT), ki se nahaja pred 3'-LTR; d) domena gag (vključuje domeno cinkovega prsta); d) zaporedja pol domene: aspartatna proteinaza (AP), reverzna transkriptaza (RT), RNaza H (RH) in integraza (IN).

Na osnovi filogenetskih analiz in strukturnih značilnosti so LTR-retroelemente uvrstili v več naddružin: Ty1/Copia (Pseudoviridae), Ty3/Gypsy (Metaviridae), Bel/Pao in Retroviridae (retrovirusi). Retroelementi brez LTR obsegajo elemente Dirs, Xena/Penelope, avtonomne elemente LINE iz naddružin L1, RTE, jockey, R2 in I ter neavtonomne elemente SINE (Wicker in sod., 2007; Kapitonov in Jurka, 2008).

#### 2.1.1 Elementi Metaviridae pri vretenčarjih

Elementi *Ty3/Gypsy* oziroma Metaviridae so LTR retroelementi, ki imajo domeno pol organizirano v naslednjem zaporedju: 5'-PR-RT-RH-IN-3'. Skupino elementov Metaviridae pri vretenčarjih tvorijo klade Kromovirus, Gmr1, Barthez, Cigr2, Mag, Osvaldo in CsRN1 (Volff in sod., 2003; Kordiš, 2005). Elementi Gmr1 so netipična klada družine Ty3/Gypsy, ki ima takšno organizacijo domene pol kot pri elementih Ty1/Copia (5'-PR-IN-RT-RH-3') (Butler in sod., 2001). Kromovirusi so elementi iz družine Metaviridae, ki imajo na 3'-koncu domene pol kromodomeno (Malik in Eickbush, 1999; Gorinšek in sod., 2004; Kordiš, 2005; Gorinšek in sod., 2005), veliko približno 50 aminokislin, največkrat pa se pojavlja pri proteinih, ki sodelujejo pri reorganizaciji kromatina (Eissenberg, 2001).

Pri sesalcih elementov Metaviridae ni (Lander in sod., 2001; Volff in sod., 2003; Gorinšek in sod. 2004; Böhne, 2008). Raznolikost in število prisotnih klad elementov Metaviridae sta pri Sauropsida (plazilci in ptiči) v primerjavi z bolj bazalnimi vretenčarji bistveno manjša (Volff in sod., 2003; Kordiš, 2009). Pri plazilcih sta poleg Kromovirusov dobro razširjeni kladi Gmr1 in Mag (Volff in sod., 2001; Goodwin in Poulter, 2002; Emerson in Thomas, 2011). Klada Mag je prisotna v genomih večine plazilcev (tuatara, kače in želve) (Kordiš, 2009). V genomu kuščarja *Anolis carolinensis* so bili identificirani številni aktivni Kromovirusi in elementi Gmr1 z močno ohranjenimi LTR-regijami (Kordiš, 2009; Emerson in Thomas, 2011), medtem ko je bila linija elementov Mag v genomu *Anolis carolinensis* izgubljena (Kordiš, 2009). Klada Mag naj bi bila izgubljena tudi pri krokodilih in ptičih. Pri krokodilih so prisotni elementi Gmr1, v genomih ptičev pa elementov Metaviridae ni, kar nakazuje, da so se elementi Metaviridae izgubili pri predniku ptičev oziroma dinozavrov – po ločitvi dinozavrov in krokodilov (Kordiš, 2009).

Genomi plazilcev se glede na prisotnost in razširjenost elementov Metaviridae (in drugih TE) bistveno razlikujejo od genomov ptičev in sesalcev. Genomi plazilcev vsebujejo številne mlade in aktivne družine TE z manjšim številom kopij na genom, in so v tem pogledu podobni nižjim vretenčarjem (Piskurek, 2009; Kordiš, 2009; Alföldi, 2011), medtem ko je raznolikost skupin TE pri ptičih in sesalcih bistveno manjša. Za ptiče in sesalce je značilno manjše število družin TE z večjim številom kopij na genom (Kordiš, 2009).

#### 2.2 VPLIV TRANSPOZICIJSKIH ELEMENTOV NA EVOLUCIJO GENOMOV

TE s svojo aktivnostjo in načinom širjenja povečujejo variabilnost znotraj genomov. Genomske spremembe lahko sprožijo na več različnih nivojih, kot sta strukturna organizacija genov in celotnih genomov ter regulacija genskega izražanja (Kidwell in Lisch, 2001; Kazazian, 2004; Biémont in Vieira, 2006; Jurka in sod., 2007). Velike in hitre genomske spremembe, ki jih TE inducirajo, so pogosto katalizator pomembnih evolucijskih novosti in sprememb (Oliver in Greene, 2009).

TE povzročajo strukturne genomske spremembe v obliki insercij, delecij, inverzij in večjih kromosomskih preureditev (Kidwell in Lisch, 2001; Kazazian, 2004). Genomsko variabilnost povečujejo z indukcijo različnih tipov homologne rekombinacije. Aktivni

in neaktivni TE so pogosti v evkariontih, kar pomeni, da so v genomih prisotna številna homologna zaporedja. Takšna zaporedja predstavljajo potencialna mesta za indukcijo homologne rekombinacije, ki povzroči podvojitve ali izgube genomskih lokusov, genov ali posameznih eksonov (Jurka, 2004). Homologi TE lahko inducirajo tudi ektopično rekombinacijo, ki povzroča večje intra- in inter-kromosomske reorganizacije (Schwartz in sod., 1998).

Strukturne genomske spremembe, ki jih inducirajo TE, pogosto pomembno vplivajo na regulacijo in stopnjo izražanja genov. TE vsebujejo (proto)promotorska, ojačevalna in utiševalna zaporedja, ki so lahko zaradi mobilne narave TE hitro razširjena po celotnem genomu, s čimer pogosto neposredno prispevajo k regulaciji genov in genskega izražanja (Feschotte, 2008). Širjenje regulatornih zaporedij z mobilnimi elementi predstavlja mehanizem, s katerim lahko postanejo ko-regulirani celotni nabori genov. Spremembe vnesene preko poškodovanja obstoječega ali dodatka novega *cis*-regulatornega elementa v regulatorno regijo tarčnega gena, vplivajo na udeleženost in vključenost tega gena v obstoječa genska regulatorna omrežja (Bourque in sod., 2008).

Vstavitev TE v promotorsko regijo vpliva na regulacijo genov prek več mehanizmov (Feschotte, 2008): a) TE v promotorsko regijo vnese novo (alternativno) mesto začetka transkripcije; b) vstavitev TE spremeni, uniči ali poškoduje obstoječe *cis*-regulatorne promotorske elemente; c) TE v promotorsko regijo vnese nove (dodatne) *cis*-regulatorne elemente; d) vstavitev TE v intronsko regijo gena lahko sproži protismerno izražanje, ki potencialno spremeni učinkovitost izražanja primarnega transkripta; e) vstavitev TE vnese nukleacijsko mesto za heterokromatin, kar povzroči utišanja tarčnega gena in sosednjih genov; e) prisotnost TE povzroči metilacijo na mestu vstavitve v genom in s tem inducira epigenetske spremembe.

Prisotnost TE lahko vpliva tudi na post-transkripcijsko uravnavanje izražanja genov (Feschotte, 2008): a) vstavitev TE v 3'-UTR regijo lahko vnese alternativno poliadenilacijsko mesto ali pa vezavno mesto za miRNA; b) TE vstavljen v intronsko regijo lahko spremeni vzorce alternativnega izrezovanja ali pa prek vnesenih novih izrezovalnih signalov v procesu eksonizacije postane eksoniziran in tako postane del spremenjenega transkripta.

#### 2.3 KAKO NASTAJAJO NOVI GENI?

Na molekularnem nivoju se organizmi na spremembe v okolju prilagajajo z adaptivno evolucijo svojih DNA-, RNA- in proteinskih zaporedij. Poleg mutacij, s katerimi se v genom organizma vnaša variabilnost, na adaptivno evolucijo in razvoj fenotipskih novosti vpliva tudi pridobivanje novih genov oziroma novih genskih struktur (Kaessmann, 2010). Nov gen lahko nastane na dva načina: prek modifikacije obstoječega gena z gensko podvojitvijo, transpozicijo in retrotranspozicijo, horizontalnim (lateralnim) prenosom, fuzijo/fizijo gena, premeščanjem eksonov (ang.

»exon shuffling«), eksonizacijo in molekularnno udomačitvijo (eksaptacijo) (Long in sod., 2003; Kaessmann, 2010). Drugi temeljni način je de novo pridobivanje novih genov znotraj naključnih medgenskih regij, intronskih regij obstoječih genov in obstoječih odprtih bralnih okvirjev (ang. »overprinting«) (Ohno, 1984; Sorek, 2007; Kaessmann, 2010).

Prevladujoči način nastajanja novih genov iz obstoječih genov so genske podvojitve, ki so značilne za evkarionte in prokarionte (Long in sod., 2003). Podvojitve kromosomskih segmentov, ki lahko vključujejo celotne gene oziroma njihove fragmente, nastanejo zaradi napak in nepravilnosti med rekombinacijo pri mejozi (Long in sod., 2003). V večjem obsegu se geni lahko podvojijo tudi ob podvojitvah celotnih genomov pri nastanku različnih poliploidij (Van de Peer in sod., 2009). Usoda novonastale kopije gena oziroma genov po genski podvojitvi je odvisna od več dejavnikov in hkrati pomeni priložnost za pridobitev nove funkcije podvojenega gena, saj izvorna kopija gena opravlja prvotno funkcijo, nova kopija pa ni pod selekcijskim pritiskom (Long in sod., 2003). V večini primerov postane novonastali gen po podvojitvi psevdogeniziran, in se iz genoma sčasoma izgubi, včasih pa pridobi novo funkcijo (neofunkcionalizacija), oziroma se funkcija izvornega gena razdeli med dve hčerinski kopiji (subfunkcionalizacija) (Kaessmann, 2010; Innan in Kondrashov, 2010). Usoda novonastale kopije gena in evolucija izvornega gena po duplikaciji sta odvisni od selekcijskega pritiska (negativna/pozitivna selekcija ali nevtralna evolucija) na genski par v času funkcionalne diverzifikacije genov (ang. »gain/loss of function«), kar lahko povzroči fiksacijo novonastalega gena. Izvorni gen lahko obdrži prvotno funkcijo, prvotna funkcija se lahko razdeli (subfunkcionalizacija), lahko pa po duplikaciji pridobi več različnih (novih) funkcij. Nova kopija lahko opravlja enako funkcijo kot izvorni gen (spremeni se število kopij gena – ang. »gene dosage«), lahko si deli del prvotne funkcije, ali pa pridobi novo funkcijo oziroma več novih funkcij (Innan in Kondrashov, 2010).

Verjetnost za nastanek novega gena iz predhodno nekodirajočega zaporedja je majhna. Predstavlja redkejši proces nastajanja novih genov kot nastanek novih genov v procesu genske duplikacije, vendar ima vseeno nezanemarljiv prispevek k evoluciji novih proteinov (Sabath in sod., 2012; Neme in Tautz, 2013). Za nastanek gena iz prej nekodirajoče genomske regije morata biti izpolnjena vsaj dva pogoja: (a) zaporedje DNA mora postati transkripcijsko aktivno; (b) znotraj transkripcijsko aktivne regije mora nastati odprti bralni okvir, ki kodira za organizem potencialno koristen protein (Kaessmann, 2010). Molekularni mehanizem de novo nastanka gena vključuje nastanek novega odprtega bralnega okvira z mutacijami, ki odstranjujejo stop-kodone in druga zaporedja, ki prekinjajo odprti bralni okvir. Novonastali odprti bralni okvir transkripcijsko aktivnost pridobi z evolucijo promotorskih elementov v 5'-regiji novega zapisa (Kaessmann, 2010). Poleg nastanka novih genov v medgenskih in intronskih regijah, novi geni pogosto nastanejo tudi znotraj zaporedij obstoječih genov, z uporabo alternativnega bralnega okvira (ang. »overprinting«) (Neme in Tautz, 2013).

Poleg mehanizma genske podvojitve prek DNA-zaporedij lahko nov gen kot kopija obstoječega gena nastane tudi preko RNA-intermediata v procesu retrotranspozicije (Brosius, 1991; Kaessmann, 2009). Med retrotranspozicijo se zapis mRNA izvornega gena reverzno prepiše, komplementarna kopija DNA (retrogen) pa se vstavi na novo mesto v genomu. Encimi, ki so potrebni za retrotranspozicijo (reverzna transkriptaza, integraza), so kodirani v različnih retrotranspozicijskih elementih, ti pa so aktivni v določeni vrsti – pri sesalcih so to elementi LINE-1 (Ensault in sod., 2000). Retrogen, vstavljen v novi genomski lokus, ohrani samo eksone izvornega starševskega gena, saj se introni izvornega gena izrežejo pri procesiranju zapisa mRNA (Kaessmann, 2009). Zaradi narave procesa retrotranspozicije retrokopija gena poleg intronov izgubi tudi promotorsko regijo, ki je izvornemu genu omogočala transkripcijo. Ključnega pomena za uspešno transkripcijo retrogenov je zato pridobivanje novih regulatornih in promotorskih zaporedij (Kaessmann, 2009).

#### 2.4 TRANSPOZICIJSKI ELEMENTI KOT VIR NOVIH KODIRAJOČIH ZAPOREDIJ

TE oziroma njihovi ostanki so pogosto vir zaporedij, iz katerih nastanejo novi geni oziroma nova kodirajoča zaporedja (eksoni) (Brosius, 1999; Zdobnov, 2005; Volff, 2006). Iz zaporedja TE, vstavljenega v intronsko regijo obstoječega gena, lahko v procesu eksonizacije nastane nov ekson. Takšen ekson, ki se pogosto alternativno izrezuje, postane del obstoječega gena, oziroma del novega, spremenjenega transkripta (Sorek, 2007). Poleg nastanka novih eksonov iz zaporedij TE, so zaporedja TE pomemben vir zaporedij za nastanek novih funkcionalnih genov (Brosius, 1999; Zdobnov, 2005; Volff, 2006; Sinzelle in sod., 2009). Proces nastajanja novih genov, katerih produkt funkcionalno ni povezan s prvotno vlogo zaporedja v genomu, se imenuje molekularna udomačitev (domestikacija) oziroma eksaptacija (Miller, 1999).

#### 2.4.1 Eksonizacija

Število eksonov in intronov v genomih ni konstantno in se spreminja med evolucijo genomov. K nastanku novih eksonov prispeva več mehanizmov: podvojitve celotnih genov ali posameznih eksonov, ekson pa lahko nastane tudi na novo z eksonizacijo (Sorek, 2007). Pogost vir zaporedij, ki so vključena v eksonizacijo, so repetitivni elementi. Dogodki in molekularni mehanizmi, ki vodijo do eksonizacije, so zelo dobro raziskani v primeru za primate specifičnih retroelementov *Alu* (Lev-Maor in sod., 2003; Singer in sod., 2004; Krull in sod., 2005; Lev-Maor in sod., 2007).

Elementi *Alu* so najpogostejša oblika repetitivnih elementov, ki jih najdemo v človeškem genomu (Lander, 2001). Pri človeku je prisotnih približno milijon kopij *Alu* 

elementov, kar predstavlja približno 11% genoma. Genomska razporeditev elementov *Alu* ni naključna, saj jih večino najdemo v regijah, ki so bogate z geni (več kot 700 000 elementov *Alu* najdemo v intronih) (Lander, 2001).

Elementi *Alu* so nastali z dimerizacijo dveh ločenih monomer *Alu* (FLAM in FRAM – ang. »fosil left/right Alu monomer«), ki sta nastala iz gena za 7SL RNA pred približno 63 milijoni let (Ullu, 1984). Elementi *Alu* so se pri primatih razširili v več zaporednih fiksacijskih valovih, v katerih so nastale tri večje skupine elementov. Podskupina elementov *Alu* J je bila aktivna v času zgodnjega razvoja primatov, podskupina *Alu* S se je najhitreje širila v času divergence antropoidov pred približno 40 milijoni let, podskupina *Alu* Y pa predstavlja najmlajšo, pri hominoidih (Hominoidae) še vedno aktivno podskupino (Quentin, 1988).

Elemente Alu oziroma dele zaporedij Alu pogosto najdemo v odprtih bralnih okvirjih (ang. »ORFs«) protein-kodirajočih genov, čeprav elementi Alu (prepisuje jih RNA polimeraza III) sami po sebi ne kodirajo proteinskih oziroma peptidnih produktov (Sorek, 2002). Pri človeku vsaj 5 % vseh eksonov, ki so podvrženi alternativnemu izrezovanju, vsebuje protein-kodirajoče regije, ki so nastale iz elementov Alu (Sorek, 2002). Velika večina (85 %) zaporedij Alu udeleženih pri eksonizaciji je v protismerni (ang. »antisense«) orientaciji glede na strukturo obstoječega gena na mestu vstavitve (Sorek, 2002). Eksonizacijo elementov Alu in posledično alternativno izrezovanje novih eksonov vodijo obstoječi ali na novo nastali (npr. zamenjava ene baze) sekvenčni motivi, ki so del zaporedij Alu in so podobni izrezovalnim (ang. »splice«) mestom, udeleženim pri alternativnem izrezovanju (Schmitz, 2011). Konsenzna zaporedja Alu vsebujejo do deset 5'- (donorskih) izrezovalnih mest in trinajst potencialnih 3'-(akceptorskih) izrezovalnih mest (Sorek, 2002). Dodatni dejavnik, ki pomembno vpliva na alternativno izrezovanje je z adenozinom bogata regija na 3'- koncu elementov Alu (vstavljenih v protismerni smeri), ki predstavlja pomembno komponento 3'izrezovalnega mesta (Sorek, 2007).

Primerjava zaporedij eksoniziranih in ne-eksoniziranih elementov *Alu* je pokazala, da je za eksonizacijo elementa *Alu* potrebno relativno malo mutacij, in da so te običajno na 5'- in 3'-izrezovalnih mestih (Schmitz, 2011). Med vstavitvijo elementa *Alu* v intronsko regijo in nastankom funkcionalnega eksona lahko mine več milijonov let. Za nastanek novega, sprva alternativno izrezovanega eksona, so potrebne mutacije in spremembe v 5'- in 3'-proto-izrezovalnih mestih, v zaporedjih elementov *Alu* (Schmitz, 2011). Z dodatnimi mutacijami in spremembami izrezovalnih mest ter drugih regulatornih regij lahko alternativno izrezovani ekson postane fiksiran in s tem stalno vključen v obstoječi transkript (Singer in sod., 2004; Krull in sod., 2005).

Eksonizacija predstavlja primer udomačitve zaporedja TE z veliko evolucijsko prednostjo (Sorek, 2007). Novi ekson je lahko znotraj obstoječega transkripta podvržen alternativnemu izrezovanju, pri čemer osnovni transkript ostane nespremenjen (prisotni

sta obe obliki transkripta) in stara oblika še naprej opravlja prvotno funkcijo, nova oblika transkripta z dodatnim eksonom pa je lahko evolucijsko testirana brez večjih posledic za organizem. Eksonizacija *Alu* elementa večinoma ne privede do nastanka spremenjenih transkriptov, ki bi predstavljali evolucijsko prednost za organizem. V nekaj primerih, ko pa eksoniziran TE za organizem predstavlja selekcijsko prednost, je prišlo do fiksacije novega eksona in s tem do pridobitve nove/spremenjene biološke funkcije (Sorek, 2007; Zarnack in sod., 2013).

#### 2.4.2 Molekularna udomačitev (domestikacija)

Proteini, kodirani znotraj TE, imajo izjemno bogat nabor funkcijskih lastnosti, ki jim omogočajo uravnavanje izrezovanja, replikacije in integracije delov DNA v gostiteljski genom (Kazazian, 2004). Tako bogat nabor lastnosti in aktivnosti je zanimiv in uporaben za gostiteljske celice, predvsem iz vidika evolucijskih inovacij. Udomačitev TE predstavlja pomembno evolucijsko inovacijo, s katero so gostiteljski genomi pridobili nove funkcionalne elemente (Miller, 1999). Udomačeni transkripti TE so pogosto pridobili nove, od prvotne funkcije neodvisne vloge (Miller, 1999; Volff, 2006). Pri integrazah, reverznih transkriptazah, transpozazah, stukturnih in plaščnih (env) proteinih, je pogosto prihajalo do neofukcionalizacij in s tem do nastanka novih genov (neogenov). Udomačeni geni predstavljajo nove in stabilne funkcionalne komponente genoma, saj med evolucijo pogosto pridobijo nove biološke vloge. Udomačeni geni so med drugim odgovorni za formiranje dolgoročnega spomina, omogočili so nastanek placente, regulirajo proliferacijo celic in apoptozo ter ščitijo naš genom pred virusnimi okužbami (Volff, 2006; Sinzelle in sod., 2009).

Zaporedja TE, iz katerih lahko nastanejo novi funkcionalni geni, se od izvornih TE razlikujejo na več načinov. Medtem ko so TE v genomu prisotni v številnih kopijah in na različnih lokusih, so udomačeni TE prisotni v genomu gostitelja v eni sami kopiji in na ortolognih položajih pri različnih organizmih (Volff, 2006). Med molekularno udomačitvijo zaporedja TE izgubijo sposobnost avtonomne (retro)transpozicije, kar je pogojeno z mutacijami v funkcionalnih mestih nujnih za mobilnost elementa (npr. LTR-regije) (Volff, 2006). Udomačeni TE ohranjajo odprti bralni okvir, ki je ohranjen pri različnih vrstah. Udomačeni TE so podvrženi negativni selekciji, saj so pogosto zaznali prebitek sinonimnih nukleotidnih zamenjav nad nesinonimnimi zamenjavami (Brandt in sod., 2005b; Volff, 2006). Evolucijska ohranjenost udomačenih TE pogosto nakazuje na neofunkcionalizacijo TE, in pridobitev novih bioloških vlog udomačenih genov.

Ključno vlogo pri transformaciji iz neaktivnega TE do funkcionalnega novega gena ima pridobivanje novih promotorskih in drugih regulatornih regij. Zaporedja genov nastalih iz neaktivnih oziroma okvarjenih TE, ki ne pridobijo promotorskih regij, se zaradi naključnih mutacij hitro psevdogenizirajo in se izgubijo iz genoma.

#### 2.5 NOVI GENI IN PRIDOBIVANJE PROMOTORSKIH ZAPOREDIJ – PRIMER RETROGENOV

Retrogeni nastanejo v procesu retropozicije ob vstavitvi procesiranega zapisa mRNA izvornega starševskega gena v nov kromosomski lokus (poglavje 2.3). Da bi retrogen postal izražen, mora pridobiti regulatorne in promotorske elemente de novo (Kaessmann, 2009). Mehanizem pridobivanja regulatornih elementov udomačenih genov in retrogenov je najverjetneje podoben, saj v obeh primerih zaporedje DNA, iz katerega lahko nastane nov gen, ne nosi lastnih regulatornih elementov, ki bi omogočili uspešno transkripcijo in s tem neofunkcionalizacijo zaporedja.

Retrogeni pogosto pridobijo regulatorna in promotorska zaporedja sosednjih genov (Kaessmann, 2009). Zmožnost transkripcije zaporedja novonastalega retrogena je lahko povečana zaradi permisivnega (odprtega) stanja kromatina na mestu vstavitve, ki je posledica transkripcijske aktivnosti sosednjih genov (Kaessmann, 2009). Retrogeni pogosto uporabijo obstoječe promotorje sosednjih genov, npr s CpG dinukleotidi bogate dvo-smerne promotorje, oziroma pridobijo oddaljene promotorje, ki prej niso bili povezani z aktivnostjo drugih genov (Fablet in sod., 2009). Vir promotorskih zaporedij retrogenov so lahko tudi zaporedja TE, ki nosijo (proto)promotorska zaporedja, vstavljena pred zaporedjem retrotranspozona (Fablet in sod., 2009). Transkripcijo retrogena lahko omogoči promotorsko zaporedje, ki nastane na novo z manjšim številom naključnih nukleotidnih zamenjav (Kaessmann, 2009), oziroma promotorsko zaporedje, ki ga retrogen ohrani znotraj originalnega transkripta starševskega gena (prisotnost alternativnega promotorskega zaporedja znotraj originalnega transkripta) (Okamura in Nakai, 2008).

Dostop do oddaljenih promotorjev in regulatornih elementov retrogenom omogočajo nove intronske/eksonske strukture, ki podaljšajo njihove 5' neprevedene regije (Fablet in sod., 2009). Podobno kot pri retrogenih tudi pri udomačenih genih novi geni vsebujejo več na novo pridobljenih intronov, ki so zbrani predvsem v 5'-neprevedeni regiji genov ter so na ta način bistveno prispevali k pridobivanju promotorskih zaporedij in regulatorni evoluciji udomačenih genov (Kordiš, 2011; Kordiš in Kokošar, 2012).

### 2.6 UDOMAČENI GENI NASTALI IZ TRANSPOZICIJSKIH ELEMENTOV

V literaturi so opisani številni primeri genov, ki so nastali iz zaporedij TE (Volff, 2006; Sinzelle in sod., 2009). Do sedaj opisani udomačeni geni izvirajo iz različnih TE, saj so nastali tako iz različnih družin DNA-transpozonov, kot retrotranspozonov. Udomačeni geni, ki so nastali iz TE, niso omejeni na posamezno skupino organizmov, saj so bili identificirani tako pri rastlinah in kvasovkah kot živalih (Volff, 2006).

#### 2.6.1 Udomačeni geni nastali iz DNA transpozonov

Izvorni gen številnih udomačenih genov, nastalih iz DNA-transpozonov, so različne transpozaze (Volff, 2006). Dobro opisan primer iz transpozaze nastalega celičnega proteina je Rag1, ki skupaj s proteinom Rag2 sodeluje pri V(D)J-rekombinaciji variabilnega dela B- in T-celičnih imunoglobulinskih receptorskih genov (Kapitonov in Jurka, 2005). Iz hAT-DNA-transpozonov so nastali udomačeni geni, ki so prisotni pri sesalcih (*Tram, Zbed4, KIAA0543, P52rIPK, Buster1-3*), rastlinah (*Daysleeper, Gary*), *Dref* pri vinski mušici (Drosophila) in *Lin-15B* pri glisti *C. elegans*. Iz transpozaz različnih družin DNA-transpozonov (Pogo, Tigger/pogo, Mariner, P, PiggyBac, Harbinger in Mutator) so nastali tudi številni drugi udomačeni geni (Volff, 2006). Biološka vloga mnogo genov, nastalih iz transpozaz, še ni znana.

#### 2.6.2 Udomačeni geni nastali iz LTR retroelementov

Zaporedja LTR-retroelementov imajo funkcijsko bogat nabor lastnosti, saj kodirajo proteine z različnimi funkcionalnimi domenami, potrebnimi za retrotranspozicijo. Domene gag, integraza, proteaza, reverzna transkripta in envelope (env) so pogost vir zaporedij, iz katerih so nastali udomačeni geni (Volff, 2006).

#### 2.6.2.1 Udomačeni geni nastali iz plaščnega proteina (Env)

Gen *env* endogenih retrovirusov kodira virusni plaščni glikoprotein Env. Env je pri retrovirusih udeležen pri vstopu virusa v celico, saj uravnava interakcijo virusa s celičnimi receptorji. Iz domene env nekaterih endogenih retrovirusov (npr. HERV-W pri človeku) so nastali geni, ki so udeleženi v razvoju oziroma morfogenezi placente (Mi in sod., 2000; Blaise in sod, 2003; Blaise in sod., 2005; Dupressoir in sod., 2011; Cornelis in sod., 2012), kot sta gena *syncytin-1* in *syncytin-2* pri človeku (Blaise in sod., 2005), ter *syncytin-A* in *syncytin-B* pri miši (Dupressoir in sod., 2005). Mišja gena *syncytin-A/B* nista ortologna genoma *syncytin-1/2* pri primatih, kar nakazuje na konvergentno evolucijo iz env-nastalih genov znotraj različnih linij placentalnih sesalcev (Dupressoir in sod. 2005).

#### 2.6.2.2 Udomačeni geni nastali iz domene gag

Zaporedje gag je prisotno pri retrotranspozonih ter pri (endogenih) retrovirusih. Domena gag je sestavljena iz treh poddomen: matriks, kapsida in nukleokapsida. Zaporedje gag je od zaporedja pol pogosto ločeno z zamikom bralnega okvirja (ang. »-1 frameshift«). Protein Gag je pri virusih udeležen pri sestavi viriona, v pakiranje virusnega genoma (RNA molekul) znotraj virusnih delcev in pri vezavi virusnih delcev na celične membrane (Kazazian, 2004; Volff, 2006).

Pri človeku so odkrili 85 genov, ki so nastali iz domene gag retrotranspozonov Ty3/Gypsy (Campillos in sod., 2006). Ti geni so nastali v vsaj petih neodvisnih

dogodkih molekularne udomačitve, katerim je sledila širitev genskih družin z genskimi podvojitvami. Iz domene gag nastali geni identificirani pri človeku so razvrščeni v 5 družin: SASpase (Asprv1), Sushi (Mart), Paraneoplastic (PNMA), SCAN in ARC (Campillos in sod., 2006).

Družino udomačenih genov Sushi (Mart) pri človeku sestavlja 12 genov - RGAG1, RGAG4, PEG10, RTL1, LDOC1, LDOC1L, FAM127A/B/C, C22orf29, ZCCHC5 in ZCCHC16, ki so prisotni pri različnih skupinah placentalnih sesalcev (Brandt in sod., 2005a; Brandt in sod., 2005b). V genomih vrečarjev sta bila identificirana 2 gena iz družine Sushi: PEG10, ki je prisoten pri vrečarjih in placentalnih sesalcih (Brandt, 2005a; Suzuki, 2007) in gen SIRH12, ki je prisoten samo pri vrečarjih (Ono in sod., 2011). Na osnovi zaporedja in strukturne ohranjenosti so ugotovili, da geni družine Sushi (Kromovirusi) evolucijsko izvirajo iz skupine Ty3/Gypsy LTR-retroelementov klade Sushi, s katerimi ohranjajo podobnost predvsem znotraj domene gag, gena PEG10 in RTL1 pa imata ohranjen tudi del regije pol (Brandt in sod., 2005). Izvor družine udomačenih genov Sushi ni bil natančno pojasnjen, saj iz filogenetske analize ni razvidno, ali je družina Sushi po izvoru monofiletska ali polifiletska, oziroma koliko dogodkov molekularne udomačitve je bilo udeleženih pri njenem nastanku (Brandt in sod., 2005a). Gen PEG10 znotraj kodirajoče regije ohranja strukturni element LTRretroelementov - zamik odprtega bralnega okvirja, z uporabo katerega se iz gena PEG10 lahko izražata dva prekrivajoča se transkripta, krajši transkript ORF1 in daljši ORF2 (Clark in sod., 2007). Večina udomačenih genov družine Sushi je na kromosomu X. Dva avtosomalna gena, PEG10 in RTL1, se izražata samo iz očetovega kromosoma, saj se nahajata v vtisnjenih genomskih regijah znotraj katerih je njuno izražanje uravnavano preko diferencialne metilacije (ang. »imprinting«) (Charlier in sod., 2001; Ono in sod., 2001).

Neodvisno od družine Sushi je iz skupine *Ty3/Gypsy* LTR-retroelementov nastala družina udomačenih genov PNMA (ang. »Paraneoplastic Ma antigens«), katere predstavniki (delno) ohranjajo strukturo domene gag ancestralnih retrotranspozonov (Schüller in sod., 2005). Pri človeku je bilo identificiranih 12 predstavnikov družine PNMA: *PNMA1, PNMA2, MOAP1, PNMA3, PNMA5, PNMA6a/b, ZCCHC12, ZCCHC18, PNMAL1, PNMAL2* in *CCDC8* (Campillos in sod., 2006). Podobno kot gen *PEG10* iz družine Sushi, gena *PNMA3* in *PNMA5* ohranjata zamik bralnega okvira za -1 znotraj kodirajočega zaporedja (Wills in sod., 2006).

Podobnost z domeno gag LTR-retrotranspozonov imata tudi gen *ARC* (ang. »activityregulated cytoskeleton associated«) (Campillos in sod., 2006) pri tetrapodih ter domena SCAN številnih genov pri sesalcih, ptičih in plazilcih (Edelstein in Collins, 2005). Domena SCAN je močno ohranjena domena, dolga 84 aminokislinskih ostankov in bogata z levcinom, ki deluje kot proteinski modul za interakcije med različnimi proteini z domeno SCAN (homo-oligomerne interakcije) in tudi z drugimi proteini (heterooligomerne interakcije) (Edelstein in Collins, 2005). Domeno SCAN pogosto najdemo na N-koncu proteinov z motivom cinkovega prsta C2H2. Pri genih s to domeno je značilna tudi domena KRAB (Sander in sod., 2003).

#### 2.6.2.3 Udomačeni geni nastali iz proteazne domene

Kot potencialni gen, nastal iz domene gag retrotranspozonov *Ty3/Gypsy*, je bil identificiran tudi gen *ASPRV1*, katerega N-končni del je podoben kapsidnemu delu domene gag (Campillos in sod., 2006). Poleg podobnosti z domeno gag protein ASPRV1 vključuje tudi aspartil-proteazno domeno, podobno proteaznim domenam LTR-retroelementov. Gen *ASPRV1* se pri človeku in miši izraža v epidermisu, kjer je odgovoren za ohranjanje pravilne teksture in vlažnosti zunanje plasti kože (Matsui in sod., 2006, Matsui in sod., 2011).

#### 2.6.2.4 Udomačeni geni nastali iz integrazne domene

V genomih različnih vretenčarjev so bili identificirani štirje geni, ki vsebujejo integrazni domeni LTR-retrotranspozonov homologno domeno. Gen *GIN1* je bil prvotno identificiran v genomih sesalcev (Lloréns in Marín, 2001), analiza različnih vretenčarskih genomov pa je pokazala da je prisoten tudi pri drugih vretenčarjih in strunarjih (Marín, 2010). Protein GIN1 vsebuje integrazno domeno, podobno integrazni domeni retrotranspozonov *Ty3/Gypsy*. Poleg *GIN1* je bil v genomih različnih vretenčarjev in strunarjev identificiran tudi njegov homolog, gen *GIN2*, vendar se je pri placentalnih sesalcih in pri Monotremata izgubil (Marín, 2010; Chalopin in sod., 2012). *GIN1* in *GIN2* ohranjata podobno gensko strukturo (položaji intronov in eksonov), kar nakazuje na njun skupen oziroma podoben evolucijski izvor. Glede na podobnost zaporedja ter ohranjenost položajev intronov in eksonov naj bi *GIN1* in *GIN2* nastala iz DNA-transpozonov družine GINGER, ki so svojo integrazno domeno predhodno pridobili iz LTR-retrotranspozonov (Marín, 2010).

V genomih sesalcev sta bila identificirana še gena *KRBA2* in *SCAND3*, z ohranjeno integrazno domeno, ki je podobna integrazni domeni elementov družine GINGER (Lloréns in sod., 2012). Gen *SCAND3* vsebuje domeno SCAN, integrazno domeno in dimerizacijsko domeno C-konca transpozaze elementov hAT. *KRBA2* vsebuje KRAB-domeno A in integrazno domeno. Predlagano je bilo, da sta *KRBA2* in *SCAND3* nastala kot posledica horizontalnega prenosa elementa GINGER2, ki vključuje KRAB-, SCAN-in integrazno domeno, in sicer iz insektov v skupnega prednika placentalnih sesalcev in vrečarjev (Lloréns in sod., 2012). Gen *NYNRIN* (*CGIN1*), še en integrazni gen pri sesalcih, je nastal iz integrazne domene endogenih retrovirusov (Marco in Marín, 2009).

## 2.7 BIOLOŠKA VLOGA UDOMAČENIH GENOV – RAZVOJ FENOTIPSKIH NOVOSTI

Ena izmed bolj intuitivnih, vendar nepopolnih definicij pojma evolucijska novost oziroma fenotipska novost je, da je evolucijska novost »vsaka novo pridobljena struktura ali biološka lastnost organizma, ki mu omogoča pridobiti novo funkcijo« (Mayr, 1960). Omenjena definicija evolucijske novosti je nepopolna iz več razlogov in med drugim implicira, da se je nova biološka lastnost razvila zaradi nove funkcije, ki jo ta ima, torej predpostavlja, da je naravna selekcija delovala v prid evoluciji nove biološke lastnosti zaradi njene nove funkcije, ki je predstavljala selekcijsko prednost (Moczek, 2009). Da bi naravna selekcija pri tem sploh lahko igrala vlogo mora delovati na neki že obstoječi variaciji biološke lastnosti oziroma funkcije, ki je dedna, in je zato ne moremo imeti za evolucijsko novost (Müller in Newman, 2005). Ena izmed razrešitev zgornjega problema je misel, o kateri je govoril že Darwin in sicer da evolucijska novost lahko nastane zaradi sekundarnih vplivov in kot stranski produkt drugih evolucijskih procesov, neodvisnih od naravne selekcije (Moczek, 2009). Eksaptacija je eden od možnih sekundarnih vplivov (Moczek, 2009), ki pravi, da so se številne nove biološke lastnosti lahko razvile zaradi razlogov, ki niso neposredno povezani s sedanjo vlogo teh bioloških oziroma fenotipskih lastnosti (Gould in Vrba, 1982). Ob upoštevanju eksaptacije, nam za nastanek bioloških novosti ni treba vpeljati naravne selekcije (Moczek, 2009). Obstoječa biološka lastnost, ki jo je že izoblikovala naravna selekcija oziroma nek nevtralni proces, lahko, če se znajde v novem, selektivnem okolju, znotraj katerega opravlja adaptivno vlogo, privede do evolucijske novosti. Novo selektivno okolje, ki obstoječo lastnost evolucijsko preoblikuje v novonastalo biološko lastnost, je za razvoj bioloških novosti prek eksaptacije tudi edini potrebni pogoj (Müller, 1990).

Udomačeni geni, nastali iz TE v procesu molekularne udomačitve oziroma eksaptacije, pogosto ohranijo funkcionalna zaporedja oziroma domene svojih izvornih elementov (Miller, 1999; Volff, 2006). Domene TE, kot so gag, integraza, proteaza in transpozaza, imajo bogat nabor funkcijskih lastnosti, ki jim omogočajo interakcije in manipulacije z biološkimi molekulami (Volff, 2006; Campillos in sod., 2006). Zaporedja TE oziroma njihovih funkcionalnih domen so med molekularno udomačitvijo pogosto pridobila nove biološke vloge (postala so neofunkcionalizirana) (pregl. 1 in 2). Število znanih bioloških vlog identificiranih udomačenih genov nakazuje, da so domene retrotranspozonov visoko prilagodljivi interakcijski moduli, prek katerih je možna interakcija med različnimi biološkimi molekulami (nukleinskimi kislinami in proteini) (Campillos et al. 2006; Volff 2006; Feschotte 2008). Novonastali geni lahko pridobijo nove biološke funkcije v procesu adaptivne evolucije oziroma prek evolucije regulatornih regij, s čimer razvijejo tkivno in/ali časovno odvisen vzorec izražanja (Volff, 2006).

Gen	Funkcija gena/proteina	Reference
PEG10	Izražen iz očetovega kromosoma. Nujen za razvoj placente.	Ono in sod., 2001;
	Čezmerno izražen pri različnih vrstah raka. Interagira z ALK1	Okabe in sod., 2003;
	(receptor za TGF- $\beta$ ) in SIAH1 (mediator apoptoze).	Ono in sod., 2006;
		Lux in sod., 2005
RGAG1	Manjša stopnja izražanja pri raku na modih.	Cheung in sod., 2010
RGAG4	Neznana.	
RTL1	Izražen iz očetovega kromosoma. Protismerno izraženi zapisi	Seitz in sod., 2003;
	miRNA iz materinega kromosoma ga negativno regulirajo.	Davis in sod., 2005;
	Nujen za tvorbo kapilar v stiku med plodom in placento.	Sekita in sod., 2008
	Izguba gena ali čezmerno izražanje povzroči odmrtje zarodka.	
ZCCHC5	Neznana.	
ZCCHC16	Neznana.	
LDOC1	Negativni regulator signalne poti NF-kB. Inducira apoptozo.	Nagasaki in sod.,
	Interagira s transkripcijskim faktrorjem MZF-1.	2003; Inoue in sod.,
		2005
LDOC1L	Neznana.	
FAM127a/b/c	Udeležen pri proliferaciji (FAM127a).	Parisi in sod., 2009
C22orf29	Neznana.	
SIRH12	Neznana.	

Preglednica 1: Poznane biološke vloge genov/proteinov iz družine Sushi Table 1: Diverse phenotypic roles of domesticated genes from the Sushi family

Pojav oziroma nastanek nekaterih udomačenih genov pri placentalnih sesalcih časovno sovpada z razvojem fenotipskih novosti kot sta placenta (Kaneko-Ishino in Ishino, 2012) in neokorteks (Oldham in sod., 2006, Takaji in sod., 2009). Iz primera udomačenih genov *PEG10* in *RTL1* iz družine Sushi, ki imata ključno vlogo pri razvoju in normalnem delovanju placente je razvidno, da so nekateri udomačeni geni ključno sooblikovali, oziroma celo omogočili razvoj pomembnih fenotipskih novosti njihovih gostiteljev (Kaneko-Ishino in Ishino, 2012). Biološka vloga oziroma funkcija je bila (delno) raziskana samo za manjše število udomačenih genov iz družine Sushi (pregl. 1) in za manjše število genov iz družine PNMA (pregl. 2).

Najbolje raziskana gena iz družine Sushi sta *PEG10* ter *RTL1*. Gena *PEG10* in *RTL1*, odkrita pri človeku (Ono in sod., 2001), oziroma pri ovci (*Ovis aries*) (Charlier in sod. 2001) sta vtisnjena gena, izražena iz očetovega kromosoma. Gensko vtisnjenje (ang. »imprinting«), je epigenetski mehanizem uravnavanja izražanja genov, pri živalih značilen za sesalce, prek katerega se gen monoalelno izraža v odvisnosti od izvora kromosoma (od očeta ali matere) (Reik in Walter, 2001). Pri človeku in miši je bilo odkritih že več kot 80 genov, katerih izražanje je podvrženo regulaciji z vtisnjenjem. Vtisnjeni geni, izraženi iz očetovega ali materinega kromosoma, so običajno znotraj vtisnjenih genomskih regij, kjer tvorijo gruče vtisnjenih genov (Kaneko-Ishino in sod., 2006). Evolucijsko je pojav regulacije izražanja genov z vtisnjenje povezan s pojavom živorodnosti (Kaneko-Ishino in Ishino, 2012). Gensko vtisnjenje je prisotno samo pri dveh živorodnih skupinah vretenčarjev – pri placentalnih sesalcih in vrečarjih–, hkrati pa so v vtisnjenih regijah geni, ki so nujno potrebni za razvoj placente in normalen razvoj zarodka (Kaneko-Ishino in Ishino, 2012). Izbitje mišjega gena *PEG10*, ki se

nahaja na kromosomu 6 v regiji skupaj s še enim genom, izraženim iz očetovega kromosoma (SGCE) in vsaj štirimi geni, izraženimi iz materinega kromosoma (PPP1R9A, PON2, PON3 in ASB4), povzroči smrt zarodka tekom zgodnjega embrionalnega razvoja ter hude morfološke okvare placente (Ono in sod., 2006). Podobno kot PEG10, se še en udomačeni gen iz družine Sushi nahaja v vtisnjeni genomski regiji kromosoma 12 pri miši, oziroma 14 pri človeku (Kaneko-Ishino in Ishino, 2012) – gen RTL1 je nujno potreben za pravilen razvoj kapilar endotelijskih celic zarodka v stiku s placento (Sekita in sod., 2008). Izbitje gena RTL1 povzroči zmanjšan pretok krvi med placento in zarodkom, kar vodi do okrnjene rasti zarodka ter odmrtja zarodka v poznem embrionalnem razvoju oziroma zgodnjem neonatalnem obdobju (Sekita in sod., 2008). Aktivnost in stabilnost molekul mRNA, ki kodirajo RTL1, uravnavajo številni transkripti RNAi, ki so iz materinega kromosoma izraženi protismerno glede na zapis RTL1 iz očetovega kromosoma (Davis in sod., 2005).

Poleg vloge pri razvoju placente je bilo za ta PEG10 dokazano, da interagira z aktivinskemu receptorju podobno kinazo 1 (ALK1), ki je TGFβ1-tip receptorja in inhibira od ALK1 in ALK5-odvisno celično signaliziranje (Lux in sod., 2005). Gen *PEG10* ima pogosto povišano izražanje pri različnih oblikah raka jeter (Okabe in sod., 2003). Za protein PEG10 je bilo ugotovljeno, da interagira s proteinom SIAH1, mediatorjem apoptoze, pri čemer povišano izražanje *PEG10* v celicah zmanjša stopnjo celične smrti, odvisne od SIAH1 (Okabe in sod., 2003). Poleg gena *PEG10* je bil s procesom uravnavanja apoptoze povezan tudi gen *LDOC1*, katerega proteinsko zaporedje vključuje SH3-vezavno domeno in levcinsko zadrgo. Gen *LDOC1* ima zmanjšano stopnjo izražanja pri nekaterih celičnih linijah raka (Nagasaki in sod., 1999) in je negativni regulator signalne poti NF-κB (Nagasaki in sod., 2003). Povečana ekspresija *LDOC1* v nekaterih celičnih linijah povzroči apoptozo in zmanjša sposobnost preživetja celic (Inoue in sod., 2005). Pokazano je bilo, da je učinkovitost LDOC1 pri sproženju apoptoze odvisna od interakcije med LDOC1 in transkripcijskim faktorjem MZF-1 v celičnem jedru (Inoue in sod., 2005).

Za tri gene iz družine udomačenih genov PNMA je bilo dokazano, da so vključeni v proces regulacije apoptoze (pregl. 2). Protein MOAP1 vsebuje BH3-podobno vezavno domeno, se veže na protein BAX in je potreben za sprožitev apoptoze, posredovane z receptorjev smrti (ang. »death receptors«) (Tan in sod., 2001). Običajno je MOAP1 v celici v obliki neaktivnega homodimera. Ob vezavi proteina RASSF1A na protein MOAP1 se inhibitorne intermolekularne interakcije prekinejo, s tem se poveča vezava MOAP1 na protein BAX, kar omogoči konformacijsko spremembo BAX in s tem apoptozo (Baksh in sod., 2005). Pri *PNMA1* mehanizem pro-apoptotičnega delovanja ni natančno pojasnjen (Chen in D'Mello, 2010). PNMA1 je ravno tako kot protein MOAP1 dimer, njegovo pro-apoptotično delovanje znotraj nevronskih celic pa je odvisno od prisotnosti BH3-podobne domene na N-koncu proteina (Chen in D'Mello, 2010). Za protein CCDC8 je bilo dokazano, da regulira aktivnost proteina P53 in je potreben za

sprožitev s P53 posredovane apoptoze pri poškodbah DNA (Dai in sod., 2011). CCDC8 interagira tako s proteinom P53 kot acetiltransferazo Tip60 in s tem regulira stopnjo acetilacije lizina na mestu 120 proteina P53, kar posledično aktivira pro-apoptotične tarče, specifične za aktivacijo pro-apoptotičnega proteina PUMA. Izguba *CCDC8* negativno regulira s strani P53 posredovano aktivacijo proteina PUMA pri poškodbah DNA (Dai in sod., 2011).

Gen	Funkcija gena/proteina	Reference
PNMA1	Nevronski pro-apoptotični protein.	Chen in D'Mello, 2010
PNMA2	Neznana.	
PNMA3	Neznana.	
MOAP1	Indukcija apoptoze. Interagira s proteini BAX, BCL2, BCL2L1 in	Tan in sod., 2001;
	RASSF1A.	Baksh in sod., 2005
PNMA5	Povečano izražanje v asociacijskih področjih neokorteksa.	Takaji in sod., 2009
PNMA6a/b	Neznana.	
ZCCHC12	Transkripcijski ko-aktivator BMP celičnega signaliziranja.	Cho in sod., 2008; Cho
	Interagira s proteinom CBP. Povezuje se z jedrnimi telesci PML.	in sod., 2009
ZCCHC18	Neznana.	
PNMAL1	Povišano izražanje pri celicah, vključenih v spermatogenezo in	Wu in sod., 2009
	oogenezo.	
PNMAL2	Neznana.	
CCDC8	Interagira s proteinom OBSL1. Mutacije v genu CCDC8	Laketa in sod., 2006;
	povzročajo sindrom M-3. CCDC8 je lokaliziran na plazemski	Hanson in sod., 2011;
	membrani in na koncih nevritov. Čezmerno izražanje CCDC8	Dai in sod., 2011
	povzroči povečanje števila nevritov na celico. Interagira s	
	proteinom P53. Izguba CCDC8 negativno regulira s strani P53	
	posredovano aktivacijo proteina PUMA pri poškodbah DNA.	
	Interagira z acetiltransferazo Tip60 in omogoča acetilacijo lizina	
	na mestu 120 proteina P53.	

Preglednica 2: Poznane biološke vloge genov/proteinov iz družine PNMA Table 2: Diverse phenotypic roles of domesticated genes from the PNMA family

*ZCCHC12* kodira protein, katerega N-končni del je homologen proteinom iz družine PNMA (Cho in sod., 2008). C-končni del proteina ZCCHC12 vsebuje jedrni lokalizacijski signal (NLS) in motiv cinkovih prstov. Dokazano je bilo, da ZCCHC12 deluje kot transkripcijski ko-aktivator signalne poti BMP (ang. »bone morphogenetic protein«) in s tem, preko interakcije s proteinoma SMAD1 in CBP modulira izražanje genov specifičnih za holinergične nevrone (Cho in sod., 2009).

Izražanje udomačenih genov je pogosto tkivno in vrstno specifično (Brandt in sod., 2005a; Takaji in sod., 2009). V primeru genov *PNMA1* in *PNMA5* iz družine PNMA je bila opažena razlika v vzorcu izražanja med glodalci in primati, saj sta *PNMA1* in *PNMA5* v možganih primatov izražena, pri miši pa ju niso zaznali (Takaji in sod., 2009). Za *PNMA5* je bilo ugotovljeno, da ima selektiven vzorec izražanja v asociacijskih področjih neokorteksa (Takaji in sod., 2009) – podobno kot skupina že odkritih genov, ki imajo specifičen vzorec izražanja v asociacijskih, senzoričnih ali motoričnih področjih neokorteksa in so zato bili predlagani kot pomemben dejavnik pri evoluciji in širjenju različnih predelov neokorteksa pri primatih (Yamamori, 2011).

Gen *ARC*, ki je podoben domeni gag retrotranspozonov se izraža v nevronskih celicah, njegova aktivnost pa je potrebna za ohranjanje plastičnosti sinaps in tvorjenje dolgoročnega spomina (Lyford in sod., 2005; Plath in sod., 2006). Izražanje gena *ARC* sprožijo nevronska aktivnost, učni procesi in različni senzorični dogodki. Translacija mRNA, ki kodira *ARC*, poteka na koncih dendritov, kjer se protein povezuje z aktinskimi filamenti in vpliva na njihovo dinamiko (Chowdhury in sod., 2006; Korb in Finkbeiner, 2011). Prek uravnavanja dinamike in rasti aktinskih filamentov protein ARC vpliva na gostoto postsinaptičnih membran, dinamiko transmembranskih receptorjev in s tem na jakost sinaptične povezave (Korb in Finkbeiner, 2011).

#### 2.8 VPLIV UDOMAČENIH GENOV NA GENSKA REGULATORNA OMREŽJA

Na osnovi genetskih analiz so ugotovili, da sta se rodova človek (*Homo*) in šimpanz (*Pan*) ločila pred približno 5 milijoni let (Chen in Li, 2001). Ljudje smo v tem času pridobili številne fenotipske značilnosti in lastnosti, po katerih se bistveno razlikujemo od šimpanzov, npr. bistveno povečanje zunanjega dela cerebralnega korteksa – neokorteksa, ki je pri človeku odgovoren za številne senzorične in kognitivne funkcije, vključno z zmožnostjo uporabe govornega komuniciranja (Hill in Walsh, 2005) Na evolucijski razvoj novih fenotipskih lastnosti je poleg evolucije kodirajočih zaporedij bistveno vplivala tudi evolucija regulatornih regij. Za proučevanje funkcionalnih razlik med možgani človeka in drugih primatov so pogosto uporabljali analizo genskega izražanja, pri čemer so se pristopi identificiranja genetskih omrežij izkazali za bolj učinkovite pri identificiranju funkcionalnih sprememb, kot pa primerjave razlik izražanja posameznih genov (Preuss in sod., 2004; Oldham in sod., 2006).

Udomačen gen *LDOC1* je bil identificiran kot centralni (ang. »hub«) gen enega izmed modulov ko-izraženih genov v cerebralnem neokorteksu človeka. Primerjava teh modulov je pokazala, da so pri človeku in šimpanzu bistveno bolj ohranjeni v evolucijsko starejših predelih cerebralnega korteksa, ne pa znotraj neokorteksa, kar je povezano s hitrim povečanjem velikosti neokorteksa med evolucijo človeka (Oldham in sod., 2006). Čeprav se gen *LDOC1* pri človeku in šimpanzu ne izraža diferencialno, pa je kromosomska inverzija gena *LDOC1* pri šimpanzu povzročila izgubo celotnega modula genov, povezanih z *LDOC1* (Oldham in sod., 2006). Prisotnost in evolucijski razvoj za človeka specifičnih modulov ko-izraženih genov in centralnih genov v teh omrežjih (npr. *LDOC1*) sta najverjetneje povezana z razvojem novih fenotipskih lastnosti pri človeku (Oldham in sod., 2006).

# 2.9 PROTEINSKE INTERAKCIJE GAG IN PROTEINOV, PODOBNIH GAG

Biološka vloga večine neofunkcionaliziranih genov, nastalih iz različnih funkcionalnih in strukturnih domen retrotranspozonov družine Ty3/Gypsy, še ni znana, oziroma je bila samo delno raziskana (poglavje 2.7). Na biološko vlogo genov oziroma kodiranih proteinov lahko pogosto sklepamo na osnovi interakcije proučevanega proteina s proteini z že znano biološko vlogo. V primeru udomačenih genov, nastalih iz gagdomene retrotranspozonov, so bile identificirane številne interakcije s proteini, udeleženimi pri apoptozi, z različnimi celičnimi receptorji in komponentami celičnega ogrodja (citoskeleta), kar je omogočilo anotacijo celičnih funkcij nekaterih udomačenih genov (Campillos in sod., 2006).

Izmed funkcionalnih domen oziroma genov, nastalih iz gag, ima domena SCAN največji nabor proteinskih interaktorjev in je prisotna pri več proteinih oziroma transkripcijskih faktorjih z motivom cinkovega prsta C2H2 (Edelstein in Collins, 2005). Proteini s to domeno so pogosto sposobni dimerizacije, interakcije z nukleinskimi kislinami in specifičnih interakcij z drugimi proteini z vključeno domeno SCAN. Neofunkcionalizacija domene SCAN iz gag-domene Gmr1 tipa elementov Ty3/Gypsy in njena uporabnost v obliki interakcijske domene transkripcijskih faktorjev je najverjetneje imela pomembne posledice za povečanje kompleksnosti genskih regulatornih omrežij pod kontrolo SCAN vsebujočih transkripcijskih faktorjev (Edelstein in Collins, 2005).

Strukturna vloga, ki jo domena gag igra v retrovirusih in endogenih retrovirusih (ERV) – pakiranje virusnega genoma v virione in tvorba virusom podobnih delcev v primeru ERV v citoplazmi celic – nakazuje na velik interakcijski potencial proteina gag in proteinov, podobnih gag. Ob določanju biološke vloge proteina ZCCHC12 iz družine PNMA, ki je nastal iz gag, je bila uspešno uporabljena ena izmed metod določanja potencialnih interakcijskih partnerjev proučevanega proteina - dvohibridni sistem kvasovke (ang. »yeast two hybrid«) (Cho in sod., 2008). Y2H je genetska tehnika, ki omogoča testiranje več kandidatnih interakcijskih molekul v kvasnih celicah (Fields in Song, 1989).

#### 2.10 EVOLUCIJSKI ODNOSI GLAVNIH SKUPIN VRETENČARJEV IN SESALCEV

Za razumevanje evolucijskih mehanizmov, dinamike in časa nastanka udomačenih genov, je ključnega pomena dobro poznavanje razširjenosti njihovih izvornih elementov ter evolucije njihovih gostiteljskih organizmov.

Udomačeni geni, nastali iz skupine *Ty3/Gypsy* LTR-retroelementov, so večinoma prisotni pri različnih skupinah vretenčarjev (Vertebrata), predvsem pa pri sesalcih.

Vretenčarji so raznolika skupina organizmov, ki se je v obdobju pred od 600 do 360 milijoni let ločila v 7 razredov (Hedges, 2009). Približno polovica še živečih vrst vretenčarjev spada v monofiletsko skupino tetrapodov (kopenski vretenčarji), ki obsega skupini Amphibia (dvoživke) in Amniota. Amnioti in dvoživke so se ločili pred približno 360 milijoni let (Anderson, 2008). Amnioti se naprej delijo na sesalce (Mammalia), Lepidosauria (kuščarji, kače in tuatara; Sphenodontia in Squamata), želve (Testudines), Crocodylia (krokodili in aligatorji) ter ptiče (Aves) (Shedlock in Edwards, 2009). Zadnji skupni prednik amniotov je živel pred približno 325 milijoni let, ko so se Synapsida (klada sesalcev) ločili od Sauropsida (Lepidosauria, Archosauria in Testudines). Skupini Lepidosauria in Archosauria sta se ločili pred približno 275 milijoni let, kladi Sphenodontia in Squamata pa pred približno 230 milijoni let, ptiči pa so se od Crocodylia ločili pred približno 220 milijoni let (Shedlock in Edwards, 2009).

Skupino še živečih sesalcev (Mammalia) sestavljajo tri klade - Prototheria, Metatheria in Eutheria (Placentalia) (Musser, 2003). Prototheria (stokovci) so najstarejša skupina sesalcev, ki se je od Theria (Metatheria in Eutheria) ločila pred približno 220 milijoni let (Madsen, 2009). Fosilni (Luo in sod., 2011) in molekularni podatki (Madsen, 2009) nakazujejo, da je prednik Eutheria (placentalni sesalci) živel pred približno 176 milijoni let. Placentalni sesalci (Eutheria) se delijo na 4 skupine – dve skupini gondvanskega (Xenarthra in Afrotheria), in dve skupini lavrazijskega izvora (Euarchontoglires in Laurasiatheria), ki so se ločile skoraj istočasno pred približno 100 milijoni let (Churakov in sod., 2009; Nishihara in sod., 2009).

## **3 MATERIAL IN METODE**

#### 3.1 MATERIAL

#### 3.1.1 Seznam kemikalij

Sigma
Serva
Sigma
Fluka
različni proizvajalci
Sigma
Sigma
Roche
Sigma
Fluka
Sigma
Sigma
<b>MBI</b> Fermentas
Promega
Sigma
Serva
Carlo Erba reagents
Kodak
Serva
Sigma
Sigma
Sigma
Sigma
Invitrogen
Roche
Merck
<b>MBI</b> Fermentas
Serva
<b>MBI</b> Fermentas
Merck
Merck
Pierce
Invitrogen
Thermo Scientific
Invitrogen

TEMED	Sigma
Triton X-100	Roche
Tris	Serva
TrypLE select	Invitrogen
Tween-20	Fluka
PEG 3350	Fluka
Pepton	Sigma
PBS	Gibco
X-α-gal	Clontech
YNB brez aminokislin	Sigma

#### 3.1.2 Uporabljeni encimi, protitelesa in proteini

DNA polimeraza Pfu	Promega
DNA polimeraza Taq	Thermo Scientific
Primarno zajčje protitelo proti SPTBN1	Santa Cruz Biotechnology
Primarno mišje protitelo proti PML	Santa Cruz Biotechnology
Primarno mišje protitelo proti V5	Invitrogen
Restriktaze BamHI in EcoRI (FastDigest)	Thermo Scientific
Sekundarna protitelesa konjugirana s HRP	Roche
Sekundarna protitelesa AlexaFluor 488 in 633	Invitrogen
Rekombinantni TGFβ1	Gibco

#### 3.1.3 Seznam uporabljenih gojišč (Y2H)

Minimalno gojišče (SDO/DDO/QDO):

•	YNB, brez aminokislin, z amonijevim sulfatom	6,9 g
•	Glukoza	20 g
•	CSM	610 - 740 mg
•	dH <sub>2</sub> O	do 1 l
•	Dodatki:	

#### Preglednica 3: Seznam dodatkov minimalnim gojiščem Table 3: List of additional components of the SDO/DDO/ODO media

Table 5. List of additional components of the SDO/DDO/QDO media			
Gojišče	Dodatki		
SDO	CSM – Trp; CSM – Leu		
DDO	CSM – Trp – Leu		
DDO + X-α-Gal + Aba	CSM – Trp – Leu; X-α-Gal 20 mg/ml (50 µl/100 mm ploščo); Aureobasidin		
	A (končna konc. 125 ng/ml)		
QDO + X-α-Gal + Aba	CSM – Ade – His – Leu – Trp; X-α-Gal 20 mg/ml (50 µl/100 mm ploščo);		
	Aureobasidin A (končna konc. 125 ng/ml)		
# Gojišče YPAD (2x) (tekoče):

•	Kvasni ekstrakt	20 g
•	Pepton	40 g
•	Glukoza	40 g
•	Adenin	120 mg
•	dH <sub>2</sub> O	do 11
•	Kanamicin	50 µg/ml

# Gojišče LBA:

•

•

kazeinski hidrolizat	10 g
Kvasni ekstrakt	5 g
NaCl	10 g
Agar	20 g
dH <sub>2</sub> O	do 11
Ampicilin	0,1 mg/ml

Gojišča za celične kulture:

•	DMEM	(Gibco)
•	EMEM	(ATCC)

- F12K (ATCC)
- Opti-MEM Invitrogen

# 3.1.4 Uporabljeni sevi kvasovk S. cerevisiae

## Preglednica 4: Uporabljeni sevi *S. cerevisiae* Table 4: *S. cerevisiae* strains used

Sev	Genotip	Reporterski marker
Y2HGold	MATa, trp1-901, leu2-3, 112,	AbA <sup>r</sup> , HIS3,
	ura3-52, his3-200, gal4∆, gal80∆,	ADE2, MEL1
	$LYS2:: GAL1_{UAS}$ -Gal1_TATA-His3,	
	GAL2 <sub>UAS</sub> –Gal2 <sub>TATA</sub> –Ade2	
	URA3 : : MEL1 <sub>UAS</sub> -Mel1 <sub>TATA</sub>	
	AUR1-C MEL1	
Y187	MATa, ura3-52, his3-200,	MEL1, LacZ
	ade2-101, trp1-901, leu2-3, 112,	
	gal4∆, gal80∆, met–,	
	URA3 : : GAL1 <sub>UAS</sub> -Gal1 <sub>TATA</sub> -LacZ,	
	MEL1	

# 3.1.5 Uporabljeni bakterijski sevi

- Escherichia coli DH5a (Genotip: F-  $\varphi$ 80*lac*Z $\Delta$ M15  $\Delta$ (*lac*ZYA-*arg*F)U169 *deo*R recA1 endA1 hsdR17( $\mathbf{r}_{k}$ ,  $\mathbf{m}_{k}^{+}$ ) phoA supE44 thi-1 gyrA96 relA1  $\lambda$ )
- Escherichia coli One Shot TOP10 (Genotip: F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ(araleu)7697 galU galKrpsL (StrR) endA1 nupG)

# 3.1.6 Uporabljene celične linije

```
Preglednica 5: Uporabljene celične linije
Table 5: list of cell cultures used
```

Organizem	Celična linija	Izvorno tkivo	Gojišče
Človek	HEK293T	ledvice	DMEM + 10% FBS
Človek	SHSY-5Y	nevroblastom	EMEM/F12K (1:1) + 10% FBS

# 3.1.7 Uporabljeni plazmidi

- pGBKT7 (Clontech)
- pGADT7 (Clontech)
- pCDNA3.1-TOPO (Invitrogen)
- pCDNA3-EGFP (Addgene, plazmid št. 13031)
- p3TP-Lux (Addgene, plazmid št. 11767)
- pRL-SV40 (Promega)

# 3.1.8 Uporabljeni začetni oligonukleotidi

# • Kloniranje PNMA5 v pGBKT7:

PNMA5\_F 5'- CATGGAGGCCGAATTCATGGCACTGACACTGCTAGAGGAT - 3'

PNMA5\_R 5'- GCAGGTCGACGGATCCCTATGTTTCCCAGGGCTTGGGATG - 3'

# • Kloniranje PNMA5 v sesalski ekspresijski vektor pCDNA3.1-TOPO

PNMA5\_F\_topo 5' - CACCATGGCACTGACACTGCTAGA - 3'

PNMA5\_R\_topo 5' - TGTTTCCCAGGGCTTGGGATGGCT - 3'

PNMA5\_R\_stop 5' - CTATGTTTCCCAGGGCTTGGGATGGCT - 3'

• Kloniranje PNMA5 v vektor pCDNA3-EGFP

# PNMA5\_GFP\_F 5'- TACCGAGCTCGGATCCATGGCACTGACACTGCTAGA - 3' PNMA5\_GFP\_R 5'- GATATCTGCAGAATTCTGTTTCCCAGGGCTTGGGATGG -3'

# 3.2 METODE

# 3.2.1 Pridobivanje podatkov

Aminokislinska in nukleotidna zaporedja udomačenih genov in njihovih izvornih retroelementov smo iskali znotraj različnih bioloških podatkovnih zbirk, ki so dostopne na spletnih strežnikih NCBI (National Center for Biotechnology Information; www.ncbi.nlm.nih.gov) in Ensembl (www.ensembl.org). Z orodjem BLAST (blast.ncbi.nlm.nih.gov) (Gertz in sod., 2006) smo preiskali podatkovne zbirke NR, EST, GSS, HTGS in WGS. Iskanje smo izvajali iterativno z različnimi zaporedji anotiranih človeških udomačenih genov, z zaporedji različnih retrotranspozonov družine Ty3/Gypsy in z njihovimi posameznimi domenami (gag, integraza, proteaza). Pri iskanju z orodjem BLAST smo uporabili standardne nastavitve. DNA-zaporedja smo prevedli z orodjem TRANSLATE (web.expasy.org/translate). Podatki o zaporedjih uporabljenih pri filogenomski in filogenetski analizi udomačenih genov, so na voljo v prilogi A.

# 3.2.2 Filogenomska analiza udomačenih genov

Dostopnost mnogih genomov vretenčarjev in sesalcev v podatkovnih zbirkah Ensembl in NCBI je omogočila identifikacijo in analizo številnih zaporedij udomačenih genov, nastalih iz retroelementov. Identificirali smo zaporedja udomačenih genov v genomih vseh štirih placentalnih nadredov (Laurasiatheria, Euarchontoglires, Afrotheria in Xenarthra), genomih vretenčarjev (Amphibia in Sauropsida) in strunarjev. Analizirali smo kromosomski položaj in organizacijo udomačenih genov, kromosomsko mobilnost udomačenih genov in gensko strukturo identificiranih genov. Podatke o strukturi genov smo pridobli iz genomskih podatkovnih baz Ensembl (izdaja 67) in NCBI Entrez (GenBank, izdaja 190.0).

# 3.2.3 Analiza ohranjenih kromosomskih položajev (analiza sintenije)

Za analizo sintenije smo uporabili podatke o položaju, dolžini in usmerjenosti (5'-3'/3'-5') sosednjih genov, ki so pred oziroma za udomačenimi geni na določenih kromosomih. Podatke smo pridobili iz podatkovnih baz Entrez in Ensembl. Dodatno smo si pomagali z vizualizacijo sinteničnih položajev z orodjem Genomicus (Muffato in sod., 2010). Če identiteta sosednjih genov ni bila nedvoumno razvidna (neanotirani geni), smo jo potrdili z orodjem BLAST.

# 3.2.4 Filogenetska analiza udomačenih genov

Pri filogenetski analizi smo uporabili vsa identificirana zaporedja udomačenih genov, nastalih iz gag, integrazne in proteazne domene retrotranspozonov. Aminokislinska zaporedja smo poravnali s programom Muscle (Edgar, 2004). Iz poravnave zaporedij smo odstranili mesta z manjkajočimi podatki (ang. »gaps«). Filogenetsko analizo smo izvedli s programoma MEGA 5.05 (Tamura in sod., 2011) in MrBayes 3.1.2 (Ronquist in Huelsenbeck, 2003). Pri filogenetski analizi smo uporabili metodo združevanja sosedov (NJ; ang. »neighbor-joining«) (Saitou in Nei, 1987), metodo največje verjetnosti (ML; ang. »maximum likelihood«) (Guindon in sod., 2005) in bayesovo metodo (Huelsenbeck in Ronquist, 2001), pri čemer smo za zunanjo referenčno skupino (ang. »outgroup«) uporabili zaporedja različnih vretenčarskih retroelementov. Testirali smo vse dostopne korekcijske modele, pri čemer so enostavni korekcijski modeli dali zanesljivejše rezultate kot kompleksni. Zanesljivost dobljenih topologij NJfilogenetskih dreves smo preverili z metodo »bootstraping«. Program MrBayes smo uporabljali prek portala CIPRES (CIPRES Science Gateway) (Miller in sod., 2010). Filogenetsko analizo smo s programom MrBayes izvajali 2x10<sup>6</sup> generacij (vzorčenje = 100), pri čemer smo uporabljali modele Poisson in JTT ob upoštevanju distribucije invariantnih mest. Bayesove posteriorne verjetnosti smo ocenili na podlagi zadnjih 10.000 dreves.

# 3.2.5 Filogenetska analiza elementov Ty3/Gypsy pri Deuterostomia

Aminokislinska zaporedja domen reverzne transkriptaze (RT) in ribonukleaze H (RNaze H) elementov *Ty3/Gypsy* (Metaviridae) smo poravnali z orodjem Muscle (Edgar, 2004). Iz poravnave zaporedij smo odstranili mesta z manjkajočimi podatki. Pri filogenetski analizi smo uporabili metodo NJ (Saitou in Nei, 1987), implementirano v programu MEGA 5.05 (Tamura in sod., 2011). Testirali smo vse dostopne korekcijske modele, pri čemer so enostavni korekcijski modeli dali zanesljivejše rezultate kot kompleksni. Zanesljivost dobljenih topologij filogenetskih dreves smo preverili z metodo »bootstraping«. Kot zunanjo referenčno skupino smo uporabili zaporedje elementa DIRS1 (AC131494).

# 3.2.6 Analiza promotorskih in drugih regulatornih zaporedij

S primerjalnimi analizami ortolognih položajev udomačenih genov smo raziskali nastanek in evolucijo njihovih regulatornih regij (5'-in 3'-UTR (neprevedene) regije, promotorska zaporedja). Zaporedja promotorskih regij udomačenih genov smo pridobili v genomski bazi podatkov Ensembl. Dolžino 3'-UTR zaporedij smo pridobili v bazi UTRdb (Grillo in sod., 2010). Podatke o izražanju mišjih in človeških udomačenih genov (ekspresijski profil) smo pridobili v bazi Unigene EST Profile Viewer (www.ncbi.nlm.nih.gov/unigene/). Prisotnost CpG-otokov znotraj promotorjev udomačenih genov smo analizirali z orodjem EpiGraph, ki je dostopno prek

genomskega brskalnika UCSC (Bock in sod., 2007). Ponavljajoča se zaporedja in transpozicijske elemente znotraj neprevedenih regij udomačenih genov smo analizirali z orodjem RepeatMasker (www.repeatmasker.org). Analizo potencialnih vezavnih mest za transkripcijske faktorje v promotorjih udomačenih genov smo izvedli z orodjem SABiosciences Champion ChiP Transcription Factor Search Portal (www.sabiosciences.com/chipqpcrsearch.php).

# 3.2.7 Metode molekulskega kloniranja

# Izolacija plazmidne DNA iz bakterijskih celic Za izolacijo plazmidne DNA iz bakterijskih celic smo uporabili kit za izolacijo DNA iz celic QIAprep spin MiniPrep kit (Qiagen).

## - Izolacija plazmidne DNA iz celic kvasovk

Za izolacijo plazmidne DNA iz celic kvasovk smo uporabili kit za izolacijo DNA iz celic QIAprep spin MiniPrep kit (Qiagen). Protokol proizvajalca smo spremenili tako, da smo celice pred dodatkom pufra N3 5 minut stresali s steklenimi kroglicami.

## - Pomnoževanje DNA z verižno reakcijo s polimerazo (PCR)

Tehniko PCR smo uporabili za pomnoževanju fragmentov DNA z namenom vstavitve te DNA v linearizirane plazmide. Pri posameznih reakcijah smo uporabili ustrezne začetne oligonukleotide, ki so našteti pod točko 3.1.8. Reakcije PCR smo izvedli po navodilih proizvajalca uporabljenih DNA polimeraz.

## - DNA elektroforeza na agaroznem gelu (AGE)

Elektroforezo na agaroznem gelu smo uporabili za ločevanje fragmentov DNA. Elektroforezo smo izvajali v pufru TAE pri napetosti 90V. Molekule DNA v gelu smo detektirali z dodatkom barvila SYBR Safe (Invitrogen).

# - Čiščenje DNA po AGE

Za čiščenje DNA iz agaroznih gelov smo uporabili kit QIAquick gel extraction kit (Qiagen). Pri delu smo sledili navodilom proizvajalca.

## - Rezanje DNA z restriktazami

Plazmidno DNA smo rezali z encimoma BamHI in EcoRI (FastDigest, Thermo Scientific). Pri delu smo sledili navodilom proizvajalca.

# - Transformacija kompetentnih bakterijskih celic

Uporabili smo kompetentne celice *E. coli* DH5α oziroma OneShot Top10 (Invitrogen). Pri delu smo sledili navodilom proizvajalca. V primeru

transformacije s plazmidno DNA smo celicam (50  $\mu$ l) dodali približno 100 ng očiščene DNA. V primeru priprave rekombinantnih molekul (In-Fusion PCR cloning kit, Clontech) smo kompetentnim celicam (50  $\mu$ l) dodali 2,5  $\mu$ l reakcijske mešanice.

# - Preverjanje nukleotidnega zaporedja

Za sekvenciranje rekombinantno pripravljenih plazmidov smo uporabili storitev podjetja MWG Operon (Nemčija). Za pripravo vzorcev smo sledili navodilom izvajalca storitve.

# - Transformacija kvasovk S. cerevisiae

Prekonočno kulturo celic (10 ml) smo prenesli v centrifugirko in centrifugirali 5 min pri 3.000 vrt./min. Po končanem centrifugiranju smo supernatant odlili in celice sprali s 5 ml sterilne dH<sub>2</sub>O. Celice smo znova centrifugirali 5 min pri 3.000 vrt./min. Supernatant smo odlili, dodali 0,5 ml 0,1 M litijevega acetata (LiAc) in suspenzijo celic prenesli v mikrocentrifugirko. Celice smo v mikrocentrifugirki centrifugirali 1 min pri 14.000 vrt./min. Supernatant smo odlili ter dodali 200  $\mu$ l 0,1 M LiAc. Po centrifugiranju (1 min pri 14.000 vrt./min) smo supernatant odlili in celice shranili na ledu.

Pripravili smo raztopino za transformacijo celic:

- 240 µl 50 % (m/v) PEG 3350
- 36 µl 1 M LiAc
- 50 μl enoverižne DNA (ssDNA) pripravljene s segrevanjem dvoverižne DNA (dsDNA) na 100 °C (2 min) in hitrim ohlajanjem na ledu.
- 34 µl transformirajoče DNA

Raztopino smo dobro premešali in prelili na kompetentne celice. Suspenzijo celic v raztopini smo inkubirali 30 min na 30 °C in nato še 20 min na 42 °C. Po inkubaciji smo celice sprali s tekočim gojiščem YPD in celice razmazali na ustrezno trdno gojišče SDO. Po štiri-dnevni inkubaciji smo posamezne kolonije s cepilno zanko precepili na sveže trdno gojišče.

# 3.2.8 Dvohibridni sistem kvasovke (Y2H)

# 3.2.8.1 Kloniranje PNMA5

Zapis za človeški gen *PNMA5* smo s PCR pomnožili iz genomske DNA celic HeLa. Pomnoženi zapis *PNMA5* smo s kitom In-fusion PCR cloning kit (Clontech) vstavili v linearizirani plazmid pGBKT7. Za linearizacijo smo uporabili encima BamHI in EcoRI (FastDigest, Thermo Scientific). Pravilnost vstavitve zapisa *PNMA5* v plazmid pGBKT7 smo preverili s sekvenciranjem plazmida.

# 3.2.8.2 Testiranje toksičnosti zapisa PNMA5-DB v celicah S. cerevisiae

Pripravljeni kvasni vektor pGBKT7-PNMA5 z vstavljenim zapisom za človeški protein PNMA5 smo transformirali v sev Y2HGold. Zapis pGBKT7-PNMA5 za celice ni toksičen, saj na kontrolnem gojišču SDO -Trp + X- $\alpha$ -Gal normalno rastejo. Izražanje zapisa PNMA5-DB v sevu Y2HGold ne sproži avto-aktivacije reporterskih genov, saj se celice ob rasti na kontrolnem gojišču niso obarvale modro, na gojišču z dodanim Aureobasidin-om A (SDO –Trp + X- $\alpha$ -Gal + AbA) pa ne rastejo.

## 3.2.8.3 Y2H

Za izvedbo Y2H smo uporabili kit Matchmaker Gold Yeast Two-Hybrid System proizvajalca Clontech. Pri izvedbi testa Y2H smo sledili navodilom proizvajalca.

Sev Y2HGold z vstavljenim vektorjem pGBKT7-PNMA5 smo v tekočem mediju (2x YPAD + kanamicin) križali s sevom Y187, v katerega je bila predhodno s transformacijo vstavljena cDNA-knjižnica pripravljena iz možganov zarodka človeka (Clontech Mate & Plate library – Human fetal brain). Kulturo diploidnih celic smo najprej nanesli na manj strogo selekcijsko gojišče (DDO/-Leu/-Trp/+X-α-Gal/+AbA). Modre kolonije, ki so zrasle na manj strogem selekcijskem gojišču, smo nanesli na bolj strogo selekcijsko gojišče (QDO/-Leu/-Trp/-His/-Ade/+X-α-Gal/+AbA). Tarčne vektorje, odgovorne za interakcijo, smo izolirali iz kvasovk in jih namnožili v celicah DH5α. Identiteto insertov cDNA smo določili s sekvenciranjem.

Morebitno avtoaktivacijo reporterskih genov seva Y2HGold s strani identificiranih zapisov cDNA – neodvisno od prisotnosti *PNMA5* – smo testirali s ko-transformacijo plazmidov pGBKT7-PNMA5/pGADT7-cDNA oziroma pGBKT7(prazen)/pGADT7-cDNA v celice Y2HGold (ang. »bait dependency test«). Rast ko-transformiranih sevov smo preverili na kontrolnem (DDO + X- $\alpha$ -Gal) in testnem (QDA + X- $\alpha$ -Gal + AbA) gojišču.

# 3.2.9 Ko-imunoprecipitacija

Zapis za človeški gen *PNMA5* smo s PCR pomnožili iz genomske DNA celic HeLa. Pomnoženi zapis *PNMA5* smo z uporabo kita pcDNA3.1 Directional TOPO Expression Kit (Invitrogen) vstavili v sesalski ekspresijski vektor pcDNA3.1. Vektor smo pripravili tako, da smo zapisu *PNMA5* dodali značko V5. Pravilnost vstavitve zapisa *PNMA5* v plazmid pcDNA3.1 smo preverili s sekvenciranjem plazmida.

Pripravljeni vektor pcDNA3.1-PNMA5-V5 smo izrazili v celicah SHSY-5Y. Za transfekcijo celic smo uporabili Lipofectamine 2000 (Invitrogen) (10  $\mu$ l Lipofectamine / 4  $\mu$ g DNA, plošča s 6 vdolbinicami). Po 72 urah smo pripravili lizate transfeciranih celic in ne-transfeciranih celic ter celic, ki so bile 24 ur inkubirane v mediju z dodanim rekombinantnim TGF $\beta$ 1 (25 ng/ml). Lizate jedrnih frakcij smo pripravili z uporabo kita

Universal Magnetic Co-IP kit (Active Motif), pri čemer smo sledili navodilom proizvajalca.

Lizate celic smo inkubirali z mišjim protitelesom proti V5 (Invitrogen) in nato dodali magnetne kroglice z vezanim proteinom G (Active Motif). Frakcijo eluiranih proteinov po imunoprecipitaciji smo razdelili na dva dela (2 x 10  $\mu$ l) in nanesli na gel NaDS-PAGE (7,5 % ali 15 %). Po končani elektroforezi smo izvedli prenos Western (nitrocelulozna membrana, 5 ur, 200 mA) v pufru Towbin (25 mM Tris, 192 mM glicin, 20 % metanol (v/v), 0,1 % SDS). Protein SPTBN1 smo zaznali z zajčjim protitelesom proti SPTBN1 (kat. št. sc-28272, Santa Cruz Biotechnology) v vzorcu ločenem na 7,5 % gelu. Protein PNMA5-V5 smo zaznali z mišjim protitelesom proti V5 (kat. št. R920-25, Invitrogen) v vzorcu ločenem na 15 % gelu.

# 3.2.10 Določanje celične lokalizacije proteina PNMA5

Zapis za človeški gen *PNMA5* smo s PCR pomnožili iz genomske DNA celic HeLa. Pomnožen zapis *PNMA5* smo s kitom In-fusion PCR cloning kit (Clontech) vstavili v lineariziran plazmid pCDNA3-EGFP. Za linearizacijo plazmida pCDNA3-EGFP smo uporabili encima BamHI in EcoRI (FastDigest, Thermo Scientific). Pravilnost vstavitve zapisa *PNMA5* v plazmid pCDNA3-EGFP smo preverili s sekvenciranjem plazmida.

Pripravljen vektor pcDNA3-PNMA5-EGFP smo izrazili v celicah HEK293T in SHSY-5Y. Celice smo gojili na krovnih steklih. Za transfekcijo celic smo uporabili Lipofectamine 2000 (Invitrogen) (10  $\mu$ l Lipofectamine / 4  $\mu$ g DNA, plošča s 6 vdolbinicami). 24 ur po transfekciji smo krovna stekla s transfeciranimi celicami prenesli na objektna stekla. Jedra celic smo označili z barvilom DAPI (ProLong Gold Antifade Reagent with DAPI, Invitrogen). Preparate celic smo opazovali pod flurescenčnim mikroskopom (eks. pri val. dolž. 488 nm).

# 3.2.11 Ko-lokalizacija proteina PNMA5 s PML jedrnimi telesci

Pripravljen vektor pcDNA3-PNMA5-EGFP smo izrazili v celicah HEK293T in SHSY-5Y. Celice smo gojili na krovnih steklih. Za transfekcijo celic smo uporabili Lipofectamine 2000 (Invitrogen) (10  $\mu$ l Lipofectamine / 4  $\mu$ g DNA, plošča s 6 vdolbinicami). 24 ur po transfekciji smo celice na krovnih steklih fiksirali (PBS + 4 % paraformaldehid, 10 min), permeabilizirali (PBS + 0,2 % Triton X-100, 5 min) in blokirali (PBS + 1,5 % FBS + 1 % goveji albumin). Celice smo čez noč inkubirali v blokirni raztopini skupaj z mišjimi primarnimi protitelesi proti PML (kat. št. sc-966, Santa Cruz Biotechnology), jih sprali s PBS in nato inkubirali s kozjimi sekundarnimi protitelesi proti mišjim imunoglobulinom AlexaFluor 633 (kat. št. A-21052, Invitrogen). Jedra celic smo označili z barvilom DAPI (ProLong Gold Antifade Reagent with DAPI, Invitrogen). Preparate celic smo opazovali pod flurescenčnim in konfokalnim mikroskopom (eks. pri val. dolž. 488 in 633 nm).

# 3.2.12 Določanje celične lokalizacije proteina SPTBN1

Celice SHSY-5Y smo gojili na krovnih steklih. Po 48 urah smo kontrolne celice, ter celice tretirane z rekombinantnim TGF $\beta$ 1 (25 ng/ml, 24 ur) fiksirali (100 % metanol, 5 min) in blokirali (PBS + 1,5 % FBS + 1 % goveji albumin). Celice smo čez noč inkubirali v blokirni raztopini skupaj z zajčjimi primarnimi protitelesi proti SPTBN1 (kat. št. sc-28272, Santa Cruz Biotechnology), jih sprali s PBS ter nato inkubirali z kozjimi sekundarnimi protitelesi proti zajčjim imunoglobulinom AlexaFluor 633 (kat. št. A-21071, Invitrogen). Jedra celic smo označili z barvilom DAPI (ProLong Gold Antifade Reagent with DAPI – Invitrogen). Preparate celic smo opazovali pod konfokalnim mikroskopom (eks. pri val. dolž. 633 nm).

## 3.2.13 Luciferazni test

Pripravljeni vektor pcDNA3.1-PNMA5 smo izrazili v celicah HEK293T. Za transfekcijo celic smo uporabili Lipofectamine 2000 (Invitrogen) (10 µl Lipofectamine / 4 µg DNA, plošča s 6 vdolbinicami). Vektor pcDNA3.1-PNMA5 smo izrazili v celicah HEK293T skupaj z reporterskim vektorjem p3TP-Lux (SMAD-odzivni promotor) in kontrolnim vektorjem pRL-SV40 (konstitutivni promotor) (razmerje p3TP-Lux/pRL-SV40 – 40:1). Celice smo 24 ur po transfekciji tretirali s TFG $\beta$ 1 (10 ng/ml, 24 ur) in pripravili celične lizate. Za pripravo lizatov in meritev aktivnosti luciferaze kresnice (ang. »Firefly«) in luciferaze iz *Renilla reniformis* smo uporabili kit Dual-Luciferase Reporter Assay System (kat. št. E1910, Promega). Aktivnost luciferaze *Renilla* smo uporabili kot interno kontrolo za normalizacijo rezultatov. Meritev aktivnosti luciferaz kresnice (Tecan Infinite M1000).

# 4 REZULTATI

# 4.1 IDENTIFIKACIJA UDOMAČENIH GENOV NASTALIH IZ RETROTRANSPOZONOV

Prisotnost, razširjenost in strukturno organizacijo genov, nastalih iz skupine Ty3/Gypsy LTR-retroelementov v genomih sesalcev, vretenčarjev in strunarjev, smo preverjali v genomskih podatkovnih bazah, prosto dostopnih na NCBI in EBI (Ensembl). Pri identifikaciji udomačenih genov v več kot 90 genomih različnih strunarjev smo uporabljali orodje BLAST (poglavje 3.2.1). Po genomskih podatkovnih bazah smo iskali z različnimi človeškimi zaporedji udomačenih genov družin Sushi in PNMA, genov, nastalih iz integrazne in proteazne domene retrotranspozonov, ter z zaporedji ohranjenih domen nekaterih retroelementov družine Ty3/Gypsy (gag, integraza, proteaza). V genomih različnih skupin placentalnih sesalcev, bazalnih sesalcev in vretenčarjev smo identificirali številna še neanotirana zaporedja udomačenih genov, ki so ortologna anotiranim zaporedjem udomačenih genov pri človeku in miši (priloga A). Identificirana zaporedja smo uvrstili na podlagi filogenetskih analiz in analiz ohranjenih kromosomskih položajev (sintenije), s čimer smo lahko razlikovali med paralognimi in ortolognimi zaporedji udomačenih genov.

# 4.2 TAKSONOMSKA RAZŠIRJENOST UDOMAČENIH GENOV

Poleg že anotiranih udomačenih genov, ki so bili identificirani pri organizmih iz placentalnega nadreda Boreoeutheria (Laurasiatheria in Euarchontoglires) (Brandt in sod., 2005a; Campillos in sod., 2006), smo v genomih placentalnih sesalcev iz nadredov Afrotheria in Xenarthra našli številne še neanotirane gene, ki kažejo podobnost z gag-, integrazno in proteazno domeno retrotranspozonov. Skupno smo v genomih placentalnih sesalcev našli 10 genov iz družine Sushi, 12 genov iz družine PNMA, 4 gene, podobne integrazni domeni retrotranspozonov, gen *ASPRV1*, ki je nastal iz proteazne domene in gen *ARC*, ki je nastal iz domene gag retrotranspozonov. Dokazali smo, da je večina udomačenih genov prisotna v genomih sesalcev iz vseh treh placentalnih nadredov, kar pomeni, da so bili prisotni in diverzificirani že v skupnem predniku placentalnih sesalcev.

Analiza razširjenosti udomačenih genov pri vrečarjih (Marsupialia) in stokovcih (Monotremata) je pokazala, da so v primerjavi s placentalnimi sesalci udomačeni geni pri vrečarjih in stokovcih bistveno manj razširjeni. V genomih vrečarjev smo identificirali samo en gen iz družine PNMA (*M-PNMA*), dva gena iz družine Sushi – *PEG10* in za vrečarje specifični gen *SIRH12* (Ono in sod., 2011), 5 integraznih genov (*GIN1, GIN2, SCAND3, KRBA2* in *NYNRIN*) ter gena *ARC* in *ASPRV1*. V trenutno edinem dostopnem genomu stokovcev – genomu kljunaša (*Ornithorhynchus anatinus*) – smo identificirali le tri udomačene gene (*ARC, ASPRV1* in *GIN1*).

Izvorna	Genska	Ime gena	Izvor -	Izvor -	Izvor -	Izvor -
domena	družina		strunarji	tetrapodi	Theria	placentalni sesalci
gag	Sushi					
		RGAG1				•
		RGAG4				•
		PEG10			•	
		RTL1				•
		LDOC1				•
		LDOC1L				•
		FAM127				•
		C22orf29				•
		ZCCHC5				•
		ZCCHC16				•
gag	PNMA					
		PNMA1				•
		PNMA2				•
		MOAP1				•
		PNMA3				•
		PNMA5				•
		PNMA6A/B				•
		ZCCHC12				•
		ZCCHC18				•
		PNMAL1				•
		PNMAL2				•
		CCDC8				•
		M-PNMA			•	
integraza						
		GIN1	•			
		GIN2	•			
		KRBA2			•	
		SCAND3			•	
		NYNRIN			•	
gag						
		ARC		•		
proteaza						
		ASPRV1			•	

Pregledn	ica 6: Pris	otnost udomač	éenih genov	znotraj	posameznih	taksonomskih	skupin
Table 6:	Taxonomi	c distribution	of diverse d	lomestic	ated genes		

Analiza udomačenih genov v genomih vretenčarjev in drugih strunarjev je ravno tako pokazala na omejeno razširjenost teh genov. Pri plazilcih, ptičih in dvoživkah so prisotni le geni *ARC*, *GIN1* in *GIN2*, medtem ko sta v genomih preostalih strunarjev prisotna samo gena *GIN1* in *GIN2*.

# 4.3 FILOGENETSKA ANALIZA UDOMAČENIH GENOV

Pri ugotavljanju evolucijskih odnosov med posameznimi geni in genskimi družinami novonastalih udomačenih genov smo uporabili metode filogenetske analize. Pri pripravi poravnav smo uporabili številna zaporedja iz organizmov iz različnih taksonomskih skupin, s čimer smo zagotovili široko taksonomsko vzorčenje, ki je ključnega pomena

za določitev izvora in časa nastanka udomačenih genov ter za ugotavljanje evolucijskih odnosov med posameznimi ortolognimi in paralognimi zaporedji. Za pripravo filogenetskih dreves smo uporabili poravnave aminokislinskih zaporedij nastalih iz gag, integrazne in proteazne domene retrotranspozonov. Zaradi nizke podobnosti (samo 30–40 %) posameznih genov znotraj multigenskih družin (Sushi in PNMA), smo v poravnave vključili samo najbolj ohranjene dele zaporedij (npr. domeno gag).

Filograme smo pridobili z metodo združevanja sosedov (NJ), metodo največje verjetnosti (ML), ter Bayesovo metode. Drevesa smo koreninili z različnimi zaporedji TE iz družine Metaviridae, oziroma z drugimi ustreznimi zaporedji.

# 4.3.1 Družina Sushi

Iz primerjave med filogrami pripravljenimi z metodama Bayes (sl. 1) in NJ (sl. 2) je razvidno, da so pri pripravi filogenetskih dreves genov družine Sushi z uporabo različnih metod posamezna odstopanja v vzorcu vejitev (topologiji drevesa), ki so posledica nizke ohranjenosti med posameznimi ortolognimi skupinami genov in relativno kratke dolžine zaporedij (230 AK), na osnovi katerih so bila filogenetska drevesa narejena. Topologija Bayesovega filogenetskega drevesa je podprta s statistično zanesljivejšimi vrednostmi, kot pri filogramu pripravljenemu po metodi NJ.

Na osnovi poravnave N-končnega dela kapsidne domene gag se tako v Bayes- kot v NJfilogramu preferenčno povezujejo paralogni geni *LDOC1* in *FAM127* (kromosom X), *RGAG4* (Xq13.1) in *LDOC1L* (22q13.31) ter *ZCCHC5* in *ZCCHC16* (kromosom X). Tako v Bayes- kot v NJ- filogramu se ortologni geni Sushi združujejo skupaj in so med seboj dobro ločeni. Vzorec vejitev znotraj posameznih skupin ortolognih genov Sushi ne nosi filogenetsko informativne ločljivosti in ne odraža dejanskih evolucijskih odnosov znotraj skupin placentalnih sesalcev.

V genomih vrečarjev sta prisotna dva gena iz genske družine Sushi. Kot je razvidno iz filogramov, se gen *PEG10* pri *Macropus eugenii* in *Sarcophilus harrisii* povezuje z ortolognimi geni *PEG10* pri placentalnih sesalcih. Gen *SIRH12*, identificiran v genomu vrečarja *M. eugenii*, se v filogramih ne povezuje z geni Sushi placentalnih sesalcev, kar potrjuje prejšnje ugotovitve, da je gen *SIRH12* specifičen za vrečarje (Ono in sod., 2011).

Kromovirusi se v filogenetskem drevesu povezujejo z geni *RTL1*, oziroma v primeru NJ-filograma, tudi z geni *PEG10*. Preferenčna povezava Kromovirusev z geni *RTL1* nakazuje, da so znotraj družine Sushi geni *RTL1* najbolj podobni izvornim retroelementom. Povezava različnih Kromovirusov z različnimi udomačenimi geni družine Sushi (primer NJ-filograma) nakazuje, da je v primeru družine Sushi prišlo vsaj do dveh neodvisnih molekularnih udomačitev, kar pomeni, da ima družina udomačenih genov Sushi parafiletski izvor.



Slika 1: Bayesova analiza evolucijskih odnosov med geni iz družine Sushi Figure 1: Bayesian phylogeny of the Sushi family of domesticated genes

## Legenda k sliki 1:

Drevo smo pripravili na osnovi poravnave zaporedij kapsidnega dela domene gag. Uporabili smo model zamenjav Poisson + G(4). Prikazane so posteriorne verjetnosti večje od 0,5. Merilo ustreza 0,1 aminokislinskim zamenjavam na mesto. Za primerjalno skupino smo uporabili kapsidni del domene gag retroelementa Cigr2 iz organizma *Ciona intestinalis*.

## Figure 1 legend:

This tree was inferred from the capsid of gag domain under a JTT+G(4) model. Only posterior probabilities larger than 0.5 are shown. The scale bar corresponds to 0.1 substitutions per site. Capsid part of the gag domain from the *Ciona intestinalis* Cigr2 element was used as an outgroup.



Slika 2: NJ-analiza evolucijskih odnosov med geni iz družine Sushi Figure 2: NJ phylogeny of the Sushi family of domesticated genes

## Legenda k sliki 2:

Uporabili smo model zamenjav »p-distance«. V primeru manjkajočih podatkov smo uporabili možnost »pairwise deletion«. Za poravnavo zaporedij smo uporabili kapsidni del domene gag. Prikazane so vrednosti zanesljivosti vejitev večje od 0,5. Merilo ustreza 0,1 aminokislinskim zamenjavam na mesto. Za primerjalno skupino smo uporabili kapsidni del domene gag retroelementa Cigr2 iz organizma *Ciona intestinalis*.

## Figure 2 legend:

The rooted neighbor-joining tree was inferred under a p-distance and pairwise deletion model from the gag domain. Bootstrap support values larger than 0.5 are shown. The scale bar corresponds to 0.1 substitutions per site. Capsid of the gag domain from the *Ciona intestinalis* Cigr2 element was used as an outgroup.

## 4.3.2 Družina PNMA

Za pripravo filogenetskih dreves smo zaporedja družine genov PNMA poravnali v delu, ki je homologen kapsidnemu delu domene gag (182 AK). Iz poravnave smo izključili zaporedja treh genov družine PNMA, ki kapsidnega dela domene gag nimajo, imajo pa ohranjeno domeno matriks (PNMAL1 in PNMAL2) oziroma samo C-končni del domene matriks (CCDC8). Primerjava med filogrami pripravljeni z različnimi metodami (Bayes, ML in NJ), ne prikazuje bistvenih razlik v vzorcu vejitev (sl. 3-5). Vzorec vejitev (topologija filogenetskega drevesa) je pri filogramu, pripravljenem z metodo Bayes, podprt s statistično zanesljivejšimi vrednostmi kot pri ML- in NJ- filogramih. Filograme smo zakoreninili z uporabo domene gag različnih TE klade Barthez. Tako v Bayes- kot v NJ-filogramih se ortologni geni PNMA združujejo in so med seboj dobro ločeni. Dobre ločljivosti nam ni uspelo dobiti le v primeru podvojenih genov ZCCHC12 in ZCCHC18, saj se ortologni geni ZCCHC12 in ZCCHC18 različnih organizmov v filogenetskem drevesu združujejo (ni vejitve na dve ločeni skupini). Vzorec vejitev znotraj posameznih skupin ortolognih genov PNMA ne nosi filogenetsko informativne ločljivosti in ne odraža dejanskih evolucijskih odnosov znotraj skupine placentalnih sesalcev.

Filogenetska analiza ima pri določitvi evolucijskih odnosov znotraj družine genov PNMA zaradi nizke podobnosti med posameznimi ortolognimi geni PNMA omejeno informativno vrednost, kar je razvidno tudi iz nizkih vrednosti zanesljivosti vejitev. Pri filogramih, pripravljenih z uporabo treh različnih metod se preferenčno združujeta gena *PNMA1* in *MOAP1*, ki sta v človeškem genomu na kromosomu 14 in sta najverjetneje nastala s podvojitvijo izvornega gena. Avtosomalni gen *PNMA2* (človeški kromosom 8) se v filogramu povezuje z genoma *ZCCHC12* in *ZCCHC18* (kromosom X), ki sta glede na lokacijo na kromosomu, ohranjenost in strukturo gena najverjetneje nastala z gensko podvojitvijo. Sestrska skupina genom *PNMA2*, *ZCCHC12* in *ZCCHC18* je skupina genov *PNMA3*, *PNMA5* in *PNMA6a/b* (na kromosomu X), ki se v filogramih združujejo in so najverjetneje nastali z genskimi podvojitvami (duplikacijami) izvornega gena vstavljenega v ta lokus (Xq28). V bližini skupine genov *PNMA3/PNMA5/PNMA6* 

najdemo več kopij procesiranih psevdogenov z visoko podobnostjo genom iz družine PNMA, kar nakazuje na visoko stopnjo genskih podvojitev udomačenih genov v tej kromosomski regiji. Skupino avtosomalnih genov *PNMAL1*, *PNMAL2* in *CCDC8* z ohranjenim delom matriks domene gag v gradnjo filograma nismo vključili. Nahajajo se skupaj na kromosomu 19 (19q13.32) in so – tako kot skupina *PNMA3/PNMA5/PNMA6* – najverjetneje nastali kot posledica genskih podvojitev na tem lokusu.

Pri vrečarjih smo identificirali en sam gen, ki je homolog genov iz družine PNMA. PNMA zapis pri vrečarjih (*M-PNMA*) se v filogramih ne povezuje z drugimi predstavniki genske družine PNMA. Ortolognih zaporedij *M-PNMA* pri placentalnih sesalcih nismo zaznali. Gen *M-PNMA* je pri vrečarjih v protismerni orientaciji znotraj introna avtosomalnega gena *LAMA3*.



Slika 3: Bayesova analiza evolucijskih odnosov med geni iz družine PNMA Figure 3: Bayesian phylogeny of the PNMA family of domesticated genes

## Legenda k sliki 3:

Drevo smo pripravili na osnovi poravnave zaporedij kapsidnega dela domene gag. Uporabili smo model zamenjav JTT + G(4). Prikazane so posteriorne verjetnosti, večje od 0,5. Merilo ustreza 0,1 aminokislinske zamenjave na mesto. Za primerjalno skupino smo uporabili kapsidni del domene gag različnih Metavirid iz klade Barthez.

## Figure 3 legend:

This tree was inferred from the capsid of gag domain under a JTT + G(4) model. Only posterior probabilities larger than 0.5 are shown. The scale bar corresponds to 0.1 substitutions per site. Capsids of the gag domain from diverse Barthez retroelements were used as an outgroup.

Kokošar J. Primerjalna genomika in analiza vlog genov sesalcev, nastalih iz retroelementov. Dokt. disertacija. Ljubljana, Univ. v Ljubljani, Biotehniška fakulteta, 2013



Slika 4: ML-analiza evolucijskih odnosov med geni iz družine PNMA Figure 4: ML phylogeny of the PNMA family of domesticated genes

## Legenda k sliki 4:

Drevo smo pripravili na osnovi poravnave zaporedij kapsidnega dela domene gag. Uporabili smo model zamenjav Poisson + G(4). Prikazane so vrednosti zanesljivosti vejitev večje od 0,5. Merilo ustreza 0,1 aminokislinske zamenjave na mesto. Za primerjalno skupino smo uporabili kapsidni del domene gag različnih Metavirid iz klade Barthez.

#### Figure 4 legend:

The maximum likelihood phylogenetic tree was inferred from the capsid of gag domain under a Poisson + G(4) model. Bootstrap support values larger than 0.5 are shown. The scale bar corresponds to 0.1 substitutions per site. Capsids of the gag domain from diverse vertebrate Barthez retroelements were used as an outgroup.



Slika 5: NJ-analiza evolucijskih odnosov med geni iz družine PNMA Figure 5: NJ phylogeny of the PNMA family of domesticated genes

## Legenda k sliki 5:

Drevo smo pripravili na osnovi poravnave zaporedij kapsidnega dela domene gag. Uporabili smo model zamenjav JTT + G(4). Prikazane so vrednosti zanesljivosti vejitev večje od 0,5. Skala ustreza 0,1 aminokislinske zamenjave na mesto. Za primerjalno skupino smo uporabili kapsidni del domene gag različnih Metavirid iz klade Barthez.

## Figure 5 legend:

The rooted neighbor-joining tree was inferred under a JTT + G(4) model from the capsid of gag domain. Bootstrap support values larger than 0.5 are shown. The scale bar corresponds to 0.1 substitutions per site. Capsids of the gag domain from diverse vertebrate Barthez retroelements were used as an outgroup.

## 4.3.3 Gen ARC

Iz poravnave domene gag ortologov gena ARC smo z metodo NJ naredili filogenetsko drevo, v katerem se zaporedja ARC združujejo v več dobro ločenih skupin (sl. 6). ARC geni pri placentalnih sesalcih tvorijo sestrsko skupino z ARC geni pri vrečarjih in stokovcih. Zaporedje pasavcu gena ARC pri (Dasypus *novemcinctus*) (AAGV03311726.1) v bazah podatkov ni celotno, saj je dostopen samo N-končni del proteina, zato v poravnavo ni bilo vključeno. ARC geni identificirani pri plazilcih in ptičih (Sauropsida) se združujejo in tvorijo sestrsko skupino z geni, identificiranimi pri placentalnih sesalcih. Gen, identificiran pri dvoživkah (AAMC01034049.1), predstavlja najstarejši gen ARC. Zaporedje AFYH012766361.1 pri latimeriji (Latimeria chalumnae) se v filogramu združuje z zaporedji ARC, vendar ne predstavlja udomačenega gena ARC, saj je pri latimeriji prisotno v več kopijah (spada v družino elementov Metaviridae). Kot primerjalno skupino smo uporabili zaporedja Metavirid iz klade Osvaldo.



## Slika 6: NJ-analiza evolucijskih odnosov med geni ARC

Drevo smo pripravili na osnovi poravnave domene gag genov ARC. Uporabili smo model zamenjav »Pdistance«. V primeru manjkajočih podatkov smo uporabili možnost »pairwise deletion«. Prikazane so vrednosti zanesljivosti vejitev večje od 0,5. Merilo ustreza 0,1 aminokislinske zamenjave na mesto. Za primerjalno skupino smo uporabili domeno gag različnih Metavirid iz klade Osvaldo.

#### Figure 6: NJ phylogeny of the ARC gene

The rooted neighbor-joining tree was inferred under a P-distance and pairwise deletion model from the gag domain. Bootstrap support values larger than 0.5 are shown. The scale bar corresponds to 0.1 substitutions per site. Gag domains from diverse vertebrate Osvaldo retroelements were used as an outgroup.

## 4.3.4 Gen ASPRV1

Filogenetsko drevo genov *ASPRV1* smo pripravili na osnovi poravnave njihove proteazne domene (sl. 7). Kot primerjalno skupino smo uporabili elemente Metaviride iz klade Cigr2. Filogenetska analiza zaporedij genov *ASPRV1* znotraj skupine placentalnih sesalcev ne poda filogenetsko informativne ločljivosti. Sestrski skupini *ASPRV1* genov placentalnih sesalcev sta skupini genov vrečarjev in stokovcev (*ASPRV1* gen pri kljunašu (*Ornithorhynchus anatinus*) predstavlja najstarejši *ASPRV1* gen).



## Slika 7: NJ-analiza evolucijskih odnosov med geni ASPRV1

Drevo smo pripravili na osnovi poravnave proteazne domene. Uporabili smo model zamenjav »P-distance«. V primeru manjkajočih podatkov smo uporabili možnost »pairwise deletion«. Prikazane so vrednosti zanesljivosti vejitev večje od 0,5. Skala ustreza 0,01 aminokislinske zamenjave na mesto. Za primerjalno skupino smo uporabili proteazno domeno elementov iz klade Cigr2.

## Figure 7: NJ phylogeny of the ASPRV1 gene

The rooted neighbor-joining tree was inferred under a P-distance and pairwise deletion model from the protease domain. Bootstrap support values larger than 0.5 are shown. The scale bar corresponds to 0.1 substitutions per site. Protease domains from diverse vertebrate Cigr2 retroelements were used as an outgroup.

## 4.3.5 Udomačeni geni nastali iz integrazne domene

Za pripravo filogenetskega drevesa integrazne družine udomačenih genov smo gene poravnali na osnovi njihove integrazne domene (sl. 8).

Najbolj razširjena predstavnika družine integraznih genov, *GIN1* in *GIN2*, se v filogenetskem drevesu združujeta z nekaterimi GIN podobnimi zaporedji identificiranimi pri organizmih *Petromyzon marinus*, *Ciona intestinalis* in *Branchiostoma floridae*. S skupino GIN-podobnih zaporedij se združujejo tudi geni DNA-transpozicijskih elementov družine GINGER, katerih integrazna domena je zelo

sorodna integrazni domeni LTR-retrotranspozonov (Marín, 2010). Paralogni geni *KRBA2* in *SCAND3* se v filogenetskem drevesu združujejo v dve dobro ločeni sestrski skupini. V tem primeru gre najverjetneje za gensko podvojitev izvornega gena v skupnem predniku vrečarjev in placentalnih sesalcev. *NYNRIN* geni se v filogenetskem drevesu preferenčno združujejo z integrazno domeno endogenih retrovirusov (ERV). Kot primerjalno skupino smo uporabili integraze različnih Metavirid. Drevo smo zakoreninili z zaporedjem Sushi ribe *Takifugu rubripes* (AF030881.2).



Slika 8: NJ-analiza evolucijskih odnosov v integrazni družini udomačenih genov Figure 8: NJ phylogeny of the integrase-derived genes

## Legenda k sliki 8:

Drevo smo pripravili na osnovi poravnave integrazne domene. Uporabili smo model zamenjav »Pdistance«. V primeru manjkajočih podatkov smo uporabili možnost »pairwise deletion«. Prikazane so vrednosti zanesljivosti vejitev večje od 0,5. Merilo ustreza 0,1 aminokislinske zamenjave na mesto. **Figure 8 legend:** 

The rooted neighbor-joining tree was inferred under a P-distance and pairwise deletion model from the integrase domain. Bootstrap support values larger than 0.5 are shown. The scale bar corresponds to 0.1 substitutions per site.

# 4.4 KDAJ IN KJE SO NASTALI UDOMAČENI GENI?

Filogenomska analiza udomačenih genov in genskih družin omogoča natančno določitev v kateri taksonomski skupini in kdaj so nastali udomačeni geni ter dinamiko njihove funkcionalne diverzifikacije. Analiza razširjenosti udomačenih genov znotraj različnih skupin strunarjev kaže na velike razlike v dinamiki nastajanja udomačenih genov. Najvišja stopnja dinamike nastajanja in funkcionalne diverzifikacije udomačenih genov je bila pri predniku placentalnih sesalcev, pri katerem je nastalo vsaj 20 različnih ortolognih genov iz genskih družin Sushi in PNMA, ki so ohranjeni pri vseh skupinah placentalnih sesalcev. Znotraj genske družine Sushi je bil v skupnem predniku placentalnih sesalcev in vrečarjev (Theria) že prisoten gen *PEG10*, medtem ko je bil iz genske družine PNMA pri predniku Theria prisoten gen *M-PNMA*. V genomu kljunaša zaporedij Sushi in PNMA ne zaznamo, kar postavlja izvor genskih družin Sushi in PNMA v čas pred približno 220 milijoni let (Madsen, 2009) ter širitev in diverzifikacijo teh genskih družin v čas pred približno 220 do približno 176 milijoni let (Luo in sod., 2011).

V primerjavi z velikim številom udomačenih zaporedij pri placentalnih sesalcih je v genomih ancestralnih sesalcev (vrečarji in stokovci) bistveno manj udomačenih genov. Pri vrečarjih so poleg *M-PNMA*, *SIRH12* in *PEG10* prisotni še geni *GIN1*, *GIN2*, *ARC*, *ASPRV1*, *KRBA2*, *SCAND3* in *NYNRIN*, med tem ko so pri Monotremata prisotni geni *GIN1*, *ARC* in *ASPRV1*. Geni *NYNRIN*, *SCAND3* in *KRBA2* so nastali v predniku Theria. Gen *ASPRV1* je nastal v predniku sesalcev (pred približno 220 do približno 250 milijoni let (Madsen, 2009)), gen *ARC* pa je nastal v predniku kopenskih vretenčarjev (pred približno 360 milijoni let (Clark, 2012)), saj je prisoten tudi pri Sauropsida in dvoživkah.

# 4.5 NUKLEOTIDNA IN AMINOKISLINSKA ZAPOREDJA UDOMAČENIH GENOV SO IZJEMNO OHRANJENA

Primerjava ortolognih zaporedij udomačenih genov kaže na visoko stopnjo ohranjenosti (pregl. 7). Primerjava aminokislinske ohranjenosti najstarejših udomačenih genov pokaže več kot 60 % identičnost med človeškim zaporedjem *GIN1* in njegovimi ortologi pri Sauropsida. V primeru gena *ARC* je AK-ohranjenost med človeškim

zaporedjem in zaporedjem identificiranim pri ptiču zebrici *Taeniopygia guttata*, 71 %. Znotraj placentalnih sesalcev AK-ohranjenost v primeru genov *GIN1* in *ARC* znaša več kot 90 %.

Aminokislinska ohranjenost ortolognih genov Sushi in PNMA je pri placentalnih sesalcih relativno visoka. (med 50 % in 80 %), v primeru najbolj ohranjenih genov (*PNMA1*, *LDOC1L*, *FAM127*) pa tudi višje od 90 %. AK-identičnost med *PNMA1* in paralognimi geni je pri človeku med 21 % (*PNMAL1*) in 58 % (*MOAP1*), ter v povprečju znaša približno 35 %.

Preglednica 7: Primerjava AK ohranjenosti med proteini pri človeku (*Homo sapiens*) in ortolognimi proteini pri vretenčarjih za nekatere udomačene gene Tabla 7: Comparison of sequence conservation between human (*Home sanisus*) domesticated genes

Table 7: Comparison of sequence conservation	between human	(Homo sapiens)	domesticated genes
and their orthologs in vertebrates			

	zebrica (T. guttata)	kljunaš ( <i>O. anatinus</i> )	tasmanski vrag (S. harrisii)	slon ( <i>L. africana</i> )	orangutan (P. abelii)
GIN1	63%	66%	74%	91%	99%
ARC	71%	76%	82%	90%	99%
ASPRV1	-	55% (frag.)	69%	79%	97%
PEG10	-	-	50%	81%	99%
PNMA1	-	-	-	94%	100%

# 4.6 ANALIZA OHRANJENOSTI KROMOSOMSKIH POLOŽAJEV UDOMAČENIH GENOV

Pri analizi ohranjenih kromosomskih položajev udomačenih genov (analiza ohranjene sintenije) smo uporabili genomske baze podatkov Entrez in Ensembl. Pridobili smo podatke o relativnih genomskih položajih genov, ki so v neposredni bližini (v 5'- ali 3'- smeri) udomačenih genov. Pri dvomljivih in/ali manjkajočih anotacijskih podatkih smo za ugotovitev homolognih odnosov med proučevanimi geni uporabili orodje BLAST. S primerjavo ohranjenih sinteničnih položajev udomačenih genov pri različnih organizmih smo lahko rekonstruirali izvorno (ancestralno) stanje kromosomskih lokusov na mestu vstavitve udomačenih genov.

Analiza ohranjene sintenije pri placentalnih sesalcih in drugih strunarjih kaže na visoko ohranjenost kodirajočih in nekodirajočih zaporedij znotraj genomskih lokusov, kjer so udomačeni geni (sl. 9; priloga B). Iz analize ohranjenih kromosomskih položajev je razvidno, da je pri predniku placentalnih sesalcev prišlo do nastanka, širitve in diverzifikacije številnih novonastalih udomačenih genov. Razvidno je, da so bili tako udomačeni geni, kot njihovi kromosomski položaji, v predniku placentalnih sesalcev že polno vzpostavljeni.



Slika 9: Shematski prikaz ohranjenosti sintenije genov družine Sushi Figure 9: Analysis of conserved synteny of the Sushi family of domesticated genes

## Legenda k sliki 9:

Analizirali smo kromosomske položaje genov družine Sushi pri različnih organizmih. Horizontalne črte predstavljajo ortologne gene. Udomačeni geni so predstavljeni z odebeljeno oranžno črto. Izvorno stanje kromosomskih položajev udomačenih genov je bilo rekonstruirano s primerjavo različnih sesalskih genomov.

## Figure 9 legend:

Chromosomal regions carrying domesticated genes from the Sushi family in the species considered in this analysis were compared, and neighbouring genes with conserved synteny were identified. Horizontal lines denote orthologous relationships. Each RDDG gene is represented in bold as a horizontal orange line on the chromosome. The ancestral states of the domesticated genes chromosomal positions were reconstructed from a comparison of syntenic positions between multiple mammalian lineages.

Analiza sinteničnih položajev omogoča nedvoumno določitev evolucijskih odnosov udomačenimi geni v primerih, ko filogenetska analiza ne omogoča jasnega med razlikovanja med paralognimi in ortolognimi sekvencami. Pri genskemu paru ZCCHC12/ZCCHC18 nam samo z uporabo filogenetskih metod analize ni uspelo določiti ortolognih odnosov med posameznimi geni, analiza sinteničnih položajev pa je omogočila nedvoumno razlikovanje med paralognimi in ortolognimi geni ZCCHC12/ZCCHC18 in hkrati dokazala prisotnost obeh genov v predniku placentalnih sesalcev. Mišji gen 2410018M08Rik/Scand3 (krom. lokacija 5G1.3) in človeški gen SCAND3 (krom. lokacija 6p22.1) sta si podobna tako po zaporedju kot na strukturnem nivoju (gen 2410018M08Rik nima SCAN in integrazne domene), vendar pa nista ortologa, saj lokus pri miši, ki vsebuje gen 2410018M08Rik/Scand3, ni v sinteniji z lokusom SCAND3 pri človeku in drugih placentalnih sesalcih. Mišji gen 2410018M08Rik/Scand3 je tako najverjetneje posledica genske podvojitve in kasnejše izgube izvornega gena SCAND3.

Združena uporaba podatkov pridobljenih s filogenetsko analizo, analizo strukture genov in analizo ohranjenih sinteničnih kromosomskih položajev, nakazuje na enoten izvor genske družine PNMA. Pri vrečarjih smo identificirali samo en gen, ki je homologen družini genov PNMA (*M-PNMA*). Gen *M-PNMA* je v protismerni orientaciji znotraj gena *LAMA3* (*Monodelphis domestica*, krom. 3). Pri placentalnih sesalcih analiza sinteničnih položajev ne pokaže prisotnosti *M-PNMA* gena znotraj intronske regije gena *LAMA3*. Najverjetnejša razlaga izvora genske družine PNMA je tako translokacija izvornega gena *PNMA* v predniku placentalnih sesalcev, temu pa je sledilo več podvojitev novonastalih genov.

## 4.7 KROMOSOMSKA MOBILNOST UDOMAČENIH GENOV

Geni iz družin PNMA in Sushi imajo zanimiv vzorec kromosomskih lokacij. Ugotovljeno je bilo, da je velik delež udomačenih genov, predvsem iz družine Sushi, na spolnem kromosomu X (Brandt in sod., 2005a; Brandt in sod., 2005b; Campillos in sod., 2006). Kljub dobremu poznavanju kromosomskih lokacij udomačenih genov dinamika njihove kromosomske mobilnosti v preteklosti ni bila podrobno raziskana.

S povezovanjem podatkov o kromosomskih položajih in evolucijskih odnosih udomačenih genov smo določili kromosomske lokacije nekaterih izvornih zaporedij PNMA in Sushi (sl. 10). Naša analiza kaže na visoko dinamiko nastajanja in mobilnosti udomačenih genov v predniku placentalnih sesalcev, kjer smo zaznali številne primere premikov udomačenih genov z avtosomov na kromosom X, ter z kromosoma X na avtosomalne kromosome. Po hitri fiksaciji nastajajočih udomačenih genov v predniku placentalnih sesalcev so njihove kromosomske lokacije znotraj placentalnih sesalcev ostale močno ohranjene, kar je razvidno iz analize ohranjene sintenije (poglavje 4.6).

Gen *M-PNMA*, najverjetnejši izvorni gen genske družine PNMA, je pri vrečarjih na avtosomalnem kromosomu (*Monodelphis domestica* – kromosom 3). Filogenomska analiza nakazuje, da je imel prvi skupni prednik placentalnih sesalcev (pred približno 176 milijoni let) več kopij *PNMA*, iz katerih so nastale vsaj tri skupine genov PNMA, prisotne pri sodobnih placentalnih sesalcih. Genski par *ZCCHC12/18* je pri človeku na kromosomu X, geni *PNMAL1/PNMAL2/CCDC8* so na kromosomu 19, kromosomskega izvora preostalih genov iz družine PNMA pa ni možno natančno določiti, saj so ti geni prisotni tako na avtosomih kot na spolnem kromosomu X, torej je bil izvorni gen te skupine genov prisoten ali na avtosomih (sl. 10/A), ali pa na spolnem kromosomu X (sl. 10/B).

*PEG10* in *RTL1* iz genske družine Sushi sta najverjetneje nastala iz izvornih retroelementov na avtosomalnih kromosomih. Geni družine Sushi, ki so ohranili samo domeno gag izvornih retroelementov, se preferenčno nahajajo na spolnem kromosomu X in so nastali iz retroelementov, ki so bili prisotni na avtosomih (sl. 10/C), ali kromosomu X (sl. 10/D).





A) in B) Shematski prikaz alternativnih hipotez premikov genov družine PNMA iz avtosomov na kromosom X ter iz kromosoma X na avtosomalne kromosome. Izvorni gen PNMA (*M-PNMA*) je pri vrečarjih na avtosomalnem kromosomu. Podvojitve izvornega elementa vodijo do nastanka avtosomalnih genov in X-vezanih genov (rdeče obarvani). Lokacija izvornega elementa PNMA v predniku placentalnih sesalcev ni znana. Izvorni gen PNMA je lahko avtosomalni (hipoteza A) oziroma X-vezan (hipoteza B). C) in D) Shematski prikaz alternativnih hipotez premikov genov družine Sushi iz avtosomov na kromosom X ter iz kromosoma X na avtosomalne kromosome. Izvorni retroelement polne velikosti je bil najverjetneje na avtosomu, saj sta tako *RTL1* kot *PEG10* avtosomalna gena. Geni nastali iz domene gag so nastali iz izvornih elementov Sushi, ki se nahajajo na avtosomih (hipoteza C) oziroma na kromosomu X (hipoteza D).

#### Figure 10: Chromosomal gene movement of domesticated genes

A) and B) Schematic representations of alternative hypotheses regarding the movement of PNMA family genes from the autosomes to the X-chromosome or from the X-chromosome to the autosomes. The therian PNMA gene progenitor is located on the autosome. A few independent originations from this progenitor may generate autosomal and X-chromosome (genes located on the X-chromosome are marked in red) gene movements. In the case of PNMA progenitor the ancestral chromosomal position is unknown and can be either autosomal (scenario under hypothesis A)) or on the X chromosome (scenario under hypothesis B)). C) and C) Schematic representation of alternative hypotheses regarding the movement of the sushi family genes from the autosomes to the X-chromosome or from the X-chromosome to autosomes. The chromoviral full-length progenitor was most probably located on the autosome, since the RTL1 and PEG10 genes are also autosomal. Numerous independent originations from the chromoviral gag progenitor may generate autosomal and X-chromosome (genes located on the X-chromosome are marked in red) movements. In the case of the chromoviral gag progenitor, the ancestral chromosome are marked in red) movements. In the case of the chromoviral gag progenitor, the ancestral chromosome are marked in red) movements. In the case of the chromoviral gag progenitor, the ancestral chromosome are marked in red) movements. In the case of the chromoviral gag progenitor, the ancestral chromosome are marked in red) movements. In the case of the chromoviral gag progenitor, the ancestral chromosome (scenario under hypothesis C)) or on the X chromosome are marked in red) movements. In the case of the chromoviral gag progenitor, the ancestral chromosome (scenario under hypothesis D)).

# 4.8 ANALIZA PROGENITORJEV UDOMAČENIH GENOV: ELEMENTI *TY3/GYPSY* (METAVIRIDAE) PRI DEUTEROSTOMIA

Za razumevanje evolucijskih mehanizmov, dinamike in časa nastanka udomačenih genov je ključnega pomena dobro poznavanje dinamike in razširjenosti njihovih izvornih TE. Vrsta TE, iz katerih so udomačeni geni nastali, v predhodnih analizah ni bila natančno določena (Brandt in sod., 2005a; Brandt in sod., 2005b; Campillos in sod., 2006; Llorens in Marin, 2001; Volff, 2006).

S filogenetsko analizo družin udomačenih genov smo določili izvorne skupine TE, iz ostankov katerih so proučevani geni nastali. Razvidno je, da so izvorni elementi večine proučevanih udomačenih genov elementi različnih linij retrotranspozonov iz družine Metaviridae (sl. 1-8).

Analiza prisotnosti elementov Metaviridae pri Deuterostomia kaže na prisotnost številnih skupin teh elementov v genomih plazilcev, elementov družine Metaviridae pa v genomih ptičev in sesalcev nismo zaznali (sl. 11; pregl. 8).

Metaviridae	Lepidosauria	Testudines	Archosauria (krokodili)	Archosauria (ptiči)	Mammalia
Kromovirus	+	*	*	Δ	Δ
Cigr2	$\Delta$	$\Delta$	+	$\Delta$	$\Delta$
Barthez	$\Delta$	+	+	$\Delta$	$\Delta$
Gmr1	+	+	+	$\Delta$	$\Delta$
new Mag	$\Delta$	+	$\Delta$	$\Delta$	$\Delta$
SURL	+	+	+	$\Delta$	$\Delta$
Mag	+	+	+	$\Delta$	$\Delta$

Preglednica 8: Razširjenost različnih skupin Metaviridae pri tetrapodih Table 8: Distribution of Metaviridae clades in Tetrapoda

+: prisotni celotni elementi,  $\Delta$ : elementi odsotni, \*: okvarjeni elementi

Prisotnost aktivnih elementov Metaviridae pri plazilcih (Sauropsida), sestrski skupini sesalcev, nakazuje na prisotnost številnih linij teh elementov v predniku amniotov (Kordiš in sod., 2006; Kordiš, 2009). V genomih različnih plazilcev (Sphenodon, Squamata, želve in krokodili) smo našli številne aktivne elemente iz klad Gmr1, Barthez in Cigr2, Kromoviruse ter tri različne linije elementov Mag (sl. 11).



Slika 11: NJ analiza elementov Metaviridae pri Deuterostomia

Drevo smo pripravili na osnovi poravnave domen RT in RNaza H. Uporabili smo model zamenjav »Pdistance«. V primeru manjkajočih podatkov smo uporabili možnost »pairwise deletion«. Prikazane so vrednosti zanesljivosti vejitev večje od 0,5. Kot primerjalno skupino smo uporabili element DIRS1 iz morskega ježka *Lytechinus* (AC131494).

#### Figure 11: NJ phylogeny of Metaviridae in Deuterostomia

The rooted neighbor-joining tree was inferred under a P-distance and pairwise deletion model from the RT and RNase H domain. Bootstrap support values larger than 0.5 are shown. The scale bar corresponds to 0.1 substitutions per site. The DIRS1 element from *Lytechinus* (AC131494) has been used to root this tree.

# 4.9 STRUKTURNA ORGANIZACIJA UDOMAČENIH GENOV, NASTALIH IZ RETROELEMENTOV

Podatke o strukturni organizaciji proučevanih udomačenih genov (sl. 12; pregl. 9) smo pridobili iz genomskih baz podatkov Ensembl in NCBI Entrez (GenBank). Analizirali smo velikost genov, kodirajočih zaporedij in neprevedenih regij, prisotnost intronov v kodirajočih in nekodirajočih regijah ter prisotnost in organizacijo posameznih proteinskih domen.

Domena gag retrotranspozonov in retrovirusev je sestavljena iz pod-domen matriks, kapsida in nukleokapsida (cinkov prst CCHC). Odprt bralni okvir, ki zapisuje domeno gag, je pri nekaterih družinah retrotranspozonov in retrovirusih od zapisov v regiji pol ločen z zamikom odprtega bralnega okvira za -1 (ang. »-1 programmed ribosomal frameshift«) (Kazazzian, 2004).

Znotraj genske družine PNMA imajo geni *PNMA1*, *PNMA2*, *MOAP1*, *PNMA5* in *PNMA6a/b* ohranjeno kapsidno domeno in domeno matriks gag proteina. Gena *PNMA3* in *M-PNMA* imata poleg teh dveh domen ohranjen tudi motiv cinkovega prsta (*CCHC-Zf*). Motiv CCHC je prisoten tudi pri genih *ZCCHC12* in *ZCCHC18*, ki imata ohranjeno kapsidno domeno, nimata pa domene matriks. *PNMAL1* in *PNMAL2* imata ohranjeno domeno matriks, medtem ko ima *CCDC8* ohranjen samo C-končni del domene matriks. Proteini družine PNMA so veliki med 279 (PNMAL2) in 535 (CCDC8) aminokislinami. Velikost zaporedja gag pri retrovirusih in retroelementih običajno znaša med 450 in 500 aminokislinami, kar sovpada z velikostjo proteinov PNMA3 (455 AK) in M-PNMA (458 AK), ki imata ohranjeno običajno (polno) strukturo domene gag (matriks, kapsida, nukleokapsida). Pri zapisih *PNMA3, PNMAL1* in *PNMAL2* je znotraj kodirajoče regije intron. Introni so pri družini PNMA znotraj neprevedenih regij (5'- in 3'-UTR regije) pogostejši kot znotraj kodirajočih zaporedij, saj v 5'-UTR regijah genov iz družine PNMA zaznamo prisotnost od enega do treh intronov (*ZCCHC12*) (Kordiš, 2011).


Slika 12: Struktura udomačenih genov nastalih iz retroelementov Figure 12: Analysis of gene structure of domesticated genes

#### Legenda k sliki 12:

A) Domene prisotne pri proučevanih udomačenih genih so predstavljene z različnimi barvami. B) Organizacija intronov in eksonov znotraj posameznih genov. Eksoni so predstavljeni kot pravokotniki in introni kot povezujoče črte. Sivi deli eksonov predstavljajo zaporedja, ki zapisujejo proteine. Oranžni deli eksonov predstavljajo neprevedena zaporedja. Proteinske domene udomačenih genov so shematsko predstavljene pod posameznim genom. a) družina PNMA; b) družina Sushi; c) iz integrazne domene nastali udomačeni geni; d) gen *NYNRIN*; e) gen *ASPRV1*; f) gen *ARC*. Merilo predstavlja 500 bp. Dolžine intronov so prikazane v pravilnem razmerju, razen v primeru dolgih intronov, katerih dolžine so označene posebej. RVP, proteaza; RT, reverzna transkriptaza; INT, integraza.

#### Figure 12 legend:

(A) Color-coded protein domains of Metaviridae retroelements and additional domains that are present in human RDDG genes. (B) Exon-intron organization of human domesticated genes. Exons are shown as boxes and introns as connecting lines. Gray regions of exons denote the protein-coding sequences, while untranslated regions are represented as orange boxes. Protein domains that are present in domesticated genes genes are shown schematically below the protein-coding sequences. a) PNMA family; b) sushi/Chromovirus family; c) integrase-derived genes; d) *NYNRIN* gene; e) retroelement protease-derived *ASPRV1* gene; f) *ARC* gene. Scale bar represents 500 bp. Distances between exons are drawn to scale except in the cases of extremely long introns. RVP, protease; RT, reverse transcriptase; INT, integrase.

Velikost proteinskih zaporedij znotraj družine Sushi je zelo različna – od več kot 1300 AK pri največjih genih (*RGAG1, RTL1*) do manj kot 150 AK pri najmanjših genih (*FAM127a/b/c*, *SIRH12*). Gen *PEG10* ima ohranjene kapsidno, nukleokapsidno in proteazno domeno. Gen *RTL1* ima poleg kapsidne domene ohranjeno tudi proteazno domeno in reverzno transkriptazo. *RGAG1* in *RGAG4* imata ohranjeno kapsidno domeno. Povečana velikost gena *RGAG1* v primerjavi z drugimi geni z ohranjeno kapsidno domeno je najverjetneje posledica številnih kratkih ponovitev, ki smo jih zaznali znotraj kodirajočega zaporedja gena *RGAG1*. Gena *ZCCHC5* in *ZCCHC16* imata poleg ohranjene kapsidne domene tudi cinkov prst CCHC. Geni *LDOC1*, *LDOC1L*, *FAM127a/b/c* in *C22orf29* imajo ohranjen samo N-končni del kapsidne domene. Intron znotraj kodirajoče regije je prisoten pri genu *RGAG1*. Pri genu *ZCCHC16* (*H. sapiens*) sta znotraj 5'-UTR regije dve nadpovprečno velika introna (323 kb in 48 kb).

Velikost proteina, ki ga kodira gen *ARC*, je 396 AK. Pri genu *ARC* sta prisotna matriksna in kapsidna dela domene gag. Znotraj 3'-neprevedene regije pri genu *ARC* najdemo dva introna.

Znano je, da *PNMA3* pri translaciji uporablja zamik bralnega okvirja za -1 (Wills in sod., 2006). Dodaten odprti bralni okvir na 3'-koncu zapisa *PNMA3* se v celicah HEK239 pri translaciji izraža z 18 % učinkovitostjo glede na osnovni zapis PNMA3 (Wills in sod., 2006). Zamik odprtega bralnega okvira za -1 je prisoten tudi pri genu *PEG10*, pri katerem se transkripti ORF1 in ORF1-ORF2 izražajo v tkivu, povezanem z embrionalnim razvojem (Clark in sod., 2007). Analize zaporedja nakazujejo na

prisotnost bralnega zamika za -1 tudi pri genu *PNMA5*, vendar ta ugotovitev ni bila eksperimentalno potrjena (Wills in sod., 2006).

Gen	Organizem	Velikost	Velikost	Introni	Introni	Struktura gena
	5	proteina	transkripta	(kod. regija)	(5'/3'-UTR)	0
PEG10	H. sapiens	708	6628	0	1/0	Cpd-Zf-RVP
RGAG1	H. sapiens	1388	5426	1	2/0	Cpd
RGAG4	H. sapiens	569	4339	0	0/1	Cpd
RTL1	H. sapiens	1358	4319	0	0/0	Cpd-RVP-RT
ZCCHC5	H. sapiens	475	2648	0	1/0	Cpd-Zf
ZCCHC16	H. sapiens	310	2958	0	2/0	Cpd-Zf
LDOC1	H. sapiens	146	1370	0	0/0	N-končni Cpd
LDOC1L	H. sapiens	239	5495	0	1/0	N-končni Cpd
FAM127a	H. sapiens	113	1236	0	0/0	N-končni Cpd
FAM127b	H. sapiens	113	1244	0	0/0	N-končni Cpd
FAM127c	H. sapiens	113	2017	0	0/0	N-končni Cpd
C22orf29	H. sapiens	364	6619	0	2/0	N-končni Cpd
SIRH12	M. eugenii	107	N/A	0	N/A	N-končni Cpd
PNMA1	H. sapiens	353	2635	0	0/0	Mtx-Cpd
PNMA2	H. sapiens	364	4950	0	2/0	Mtx-Cpd
PNMA3	H. sapiens	455	3630	1	1/0	Mtx-Cpd-Zf
MOAP1	H. sapiens	351	2441	0	2/0	Mtx-Cpd
PNMA5	H. sapiens	448	3197	0	1-3/0	Mtx-Cpd
PNMA6a/b	H. sapiens	399	2236	0	1/0	Mtx-Cpd
ZCCHC12	H. sapiens	402	2231	0	3/0	Cpd-Zf
ZCCHC18	H. sapiens	403	3045	0	2/0	Cpd-Zf
PNMAL1	H. sapiens	439	3689	1	1/0	Mtx
PNMAL2	H. sapiens	279	1726	1	0/0	Mtx
CCDC8	H. sapiens	538	3213	0	0/0	C-končni Mtx
M-PNMA	M. eugenii	458	N/A	0	N/A	Mtx-Cpd-Zf
ARC	H. sapiens	396	2943	0	0/2	Mtx-Cpd
ASPRV1	H. sapiens	343	2172	0	0/0	RVP
GIN1	H. sapiens	522	3229	6	1/0	Zf-INT
GIN2	G. gallus	422	1809	7	?/0	INT
NYNRIN	H. sapiens	1898	7857	7	1/0	NYN-INT
KRBA2	H. sapiens	492	1964	1	0/0	KRAB-INT
SCAND3	H. sapiens	1325	4877	3	0/0	SCAN-INT-
						hAT

Preglednica 9: Struktura udomačenih genov Table 9: Analysis of gene structure of domesticated genes

Cpd: kapsidna domena, Zf: cinkov prst, RVP: aspartatna proteaza retrovirusev, RT: reverzna transkriptaza, Mtx: domena matriks, INT: integraza, NYN: domena NYN, KRAB: domena "Kruppel associated box", SCAN: domena SCAN, hAT: dimerizacijska domena družine hAT

Evolucijsko najstarejša gena znotraj družine udomačenih genov s prisotno integrazno domeno, *GIN1* in *GIN2*, imata znotraj kodirajoče regije večje število intronov. *GIN1*, ki ima poleg integrazne domene tudi cinkov prst C2H2, znotraj kodirajoče regije vsebuje 6 intronov, medtem ko jih *GIN2* vsebuje 7. Pri genu *SCAND3* so v kodirajoči regiji trije

introni. Ekson, z domeno SCAN, je od integrazne domene ločen z intronom. Domena hAT gena *SCAND3* je ravno tako na eksonu, ločenem od integrazne domene. V primeru gena *KRBA2* sta zaporedje *KRAB* in integrazna domena ločeni z intronom. Gen *NYNRIN* ima znotraj kodirajoče regije 7 intronov in poleg integrazne domene vsebuje tudi domeno NYN. Integrazna domena gena *NYNRIN* je podobna integrazni domeni endogenih retrovirusov (ERV) (Marco in Marín, 2009).

Znotraj vretenčarskih genomov smo zaznali en gen s proteazno domeno, podobno proteaznim domenam Metavirid. Gen *ASPRV1* znotraj kodirajoče oziroma 5'/3'- neprevedene (UTR-) regije nima prisotnih intronov.

## 4.10 DE NOVO PRIDOBIVANJE IN EVOLUCIJA PROMOTORSKIH IN DRUGIH REGULATORNIH REGIJ UDOMAČENIH GENOV

Posamezni deli retrotranspozonov, npr. domene gag, proteaza in integraza, nimajo lastnih promotorskih regij. Novi geni nastali iz ostankov retrotranspozonov so morali, da bi postali izraženi, promotorske regije pridobiti na novo. Z metodami primerjalne genomike smo proučevali izvor in evolucijo na novo pridobljenih promotorskih ter 5'in 3'-neprevedenih regij udomačenih genov. Ugotovili smo, da udomačeni geni svojih promotorskih regij niso pridobili iz ostankov LTR-zaporedij in da so mehanizmi pridobivanja promotorskih regij udomačenih genov podobni mehanizmom pridobivanja promotorskih regij retrogenov (Kaessmann, 2009).

### 4.10.1 Mehanizmi neofunkcionalizacije

S filogenomsko analizo izvora, razširjenosti in strukture udomačenih genov v genomih sesalcev in drugih vretenčarjev nam je uspelo določiti ključne mehanizme pri neofunkcionalizaciji zaporedij novonastalih genov (sl. 13). Časovna opredelitev nastanka udomačenih genov (poglavje 4.4) je pokazala, da je bila večina proučevanih udomačenih genov fiksirana v predniku placentalnih sesalcev (poglavje 4.6; priloga B). Ortologna zaporedja udomačenih genov so znotraj placentalnih sesalcev dobro ohranjena (poglavje 4.5) in podvržena močni negativni selekciji (Brandt in sod., 2005b). Čas od izgube izvornih TE v predniku sesalcev (pred približno 220 do 250 milijoni let) do nastanka neofunkcionaliziranih novih genov v predniku placentalnih sesalcev (pred približno 176 milijoni let) je pomemben za razumevanje nastanka teh genov, saj nakazuje, da je bilo za nastanek udomačenih genov potrebno daljše časovno obdobje (približno 50–80 milijonov let), znotraj katerega se je zgodilo več nukleotidnih zamenjav, ki so omogočile eksonizacijo in neofunkcionalizacijo zaporedij, nastalih iz retroelementov.

Eden od ključnih korakov v procesu neofunkcionalizacije udomačenih genov je bila eksonizacija različnih domen retroelementov (gag, integraza, proteaza), iz katerih so novi geni nastali. Ostanki retroelementov, ki ne pridobijo regulatornih in promotorskih zaporedij, običajno postanejo psevdogeni, zato je pri preživetju eksoniziranih zaporedij v genomih gostiteljev pridobivanje promotorskih in drugih regulatornih zaporedij imelo ključno vlogo. Pridobljena regulatorna zaporedja novonastalim genom omogočijo dovolj visok in (pogosto) tkivno specifičen nivo izražanja, povezanost v genska regulatorna omrežja, pridobivanje novih bioloških vlog in s tem fiksacijo v genomih gostiteljev.



#### Slika 13: Ključni mehanizmi vključeni v neofunkcionalizacijo novonastalih udomačenih genov

Številne nukleotidne zamenjave omogočijo eksonizacijo ostankov retroelementov. Eksonizirana zaporedja lahko pridobijo promotorska in druga regulatorna zaporedja prek de novo evolucije regulatornih zaporedij, prek pridobivanja 5//3'-neprevedenih genskih regij, z evolucijo genske strukture in s pridobitvijo oziroma uporabo obstoječih regulatornih zaporedij, npr. dvosmernih CpG-promotorjev sosednjih genov. Na shemi so eksoni predstavljeni kot sivi pravokotniki. Z oranžno barvo so predstavljena 5'/3'-neprevedena zaporedja. Regulatorna regija, pridobljena de novo, je predstavljena z modro barvo.

#### Figure 13: Mechanisms involved in the process of neofunctionalization of domesticated genes

In the transition phase from retroelement remains to the first RDDGs, many nucleotide changes were necessary for the neofunctionalization. One of the crucial steps in the process of neofunctionalization was the exonization of retroelement domains (gag, protease and integrase), which produced ready-to-use modules. To become expressed at a significant level and in the tissues where it can exert a selectively beneficial function, a new gene needs to acquire a core promoter and other structural elements that regulate its expression. Exons and introns are shown as orange (5' and 3' UTR regions) or gray (coding part of the exons) boxes and connecting lines. A de novo acquired promoter is shown in blue.

#### 4.10.2 De novo pridobivanje in evolucija 5' UTR / promotorske regije

Analiza kromosomskih položajev in ohranjenih sintenij udomačenih genov je pokazala, da sta se fiksacije ostankov retroelementov in s tem nastanek udomačenih genov preferenčno dogajala v genomskih regijah z višjim transkripcijskim potencialom (CpG-bogate regije), ter v primerih, ko so novonastali geni v neposredni bližini obstoječih genov. Prisotnost sosednjih genov nakazuje na možnost, da lahko novonastali geni uporabijo obstoječe promotorske regije in sposobnost izražanja sosednjih genov za povečanje lastne učinkovitosti izražanja, saj fizična bližina sosednjih genov vpliva na stopnjo odprtosti kromatina in dostopnost različnih regulatornih elementov.



#### Slika 14: Različni tipi promotorjev pri udomačenih genih

A) De novo pridobitev promotorske regije. B) Pridobitev CpG-promotorja prek evolucije CpG protopromotorske regije. C) Pridobitev dvosmernega CpG-promotorja, ki je že del promotorja obstoječega gena. D) Evolucija intronsko/eksonske strukture 5'-UTR regij – pridobitev oddaljenega CpG promotorja.
E) Pridobitev obstoječega enosmernega CpG-promotorja in eksonov 5'-UTR regije – skupna uporaba promotorske regije.

#### Figure 14: Types of promoters of domesticated genes

A) Recruitment of proto-promoters from the CpG island-less region. B) Recruitment of proto-promoters from the CpG-rich island. C) Recruitment of a bidirectional (CpG-enriched) promoter from neighboring gene in the vicinity of the RDDG gene. D) Recruitment of distant promoters in the genomic neighbourhood via the acquisition of a new 5' untranslated exon–intron structure. E) Sharing of the unidirectional (CpG-enriched) promoter from a neighboring gene in the vicinity of the RDDG gene.

Izvorni	Ime gena	5'-UTR	CpG-otok/	Dvosmerni	CpG
transpozicijski	-	introni	Proto-	CpG-	otoka
element			promotor	promotor	ni
Kromovirus					
	RGAG1	Da			•
	RGAG4	Ne	•		
	PEG10	Da		•	
	RTL1	Ne			•
	LDOC1	Ne	•		
	LDOC1L	Da	•		
	FAM127A/B/C	Ne	•		
	C22orf29	Da	•		
	ZCCHC5	Da			•
	ZCCHC16	Da	•		
Barthez					
	PNMA1	Ne	•		
	PNMA2	Da		٠	
	MOAP1	Da		٠	
	PNMA3	Da	•		
	PNMA5	Da			•
	PNMA6A/B	Da	•		
	ZCCHC12	Da	•		
	ZCCHC18	Da	•		
	PNMAL1	Da	•		
	PNMAL2	Ne	•		
	CCDC8	Ne	•		
Gmr1					
	GIN1	Da		•	
	GIN2	Ne		•	
	KRBA2	Da			•
	SCAND3	Ne			•
Osvaldo					
	ARC	Ne	•		
Cigr2					
	ASPRV1	Ne			•
ERV					
	NYNRIN	Da	•		

Preglednica 10: Analiza	promotorskih	regij udo	omačenih genov
Table 10. Analysis of do	mesticated ger	es nroma	oters

Primerjalna analiza promotorskih regij udomačenih genov kaže na več različnih virov oziroma načinov pridobivanja promotorskih zaporedij teh genov. Večina (19 genov, 68 %) udomačenih genov je v procesu neofunkcionalizacije pridobila CpG-promotor, ki pred tem ni bil del promotorske regije sosednjih genov. Manjše število udomačenih genov (5 genov, 18 %) je pridobilo že obstoječ funkcionalen promotor, saj za izražanje uporabljajo dvosmerni CpG-promotor, ki si ga delijo z obratno orientiranim sosednjim genom (ang. »head to head gene pair«) (sl. 14/C). V sedmih primerih (25 %) so

udomačeni geni v procesu pridobivanja promotorske regije znotraj 5'-neprevedenih zaporedij pridobili introne. Evolucija intronsko/eksonske strukture 5'-UTR regij lahko poveča nivo izražanja genov, saj omogoča povezavo med kodirajočim zaporedjem in oddaljenimi CpG-otoki ali drugimi regulatornimi regijami, ob tem pa se zmanjšanja velikost neprevedenih eksonov (Kordiš, 2011; Kordiš in Kokošar, 2012). Ekstremen primer, ko so znotraj 5'-neprevedene regije izjemno veliki introni, je človeški gen *ZCCHC16*, pri katerem velikosti dveh intronskih regij znotraj 5'-UTR znašata približno 48 kb in približno 323 kb (sl. 14/D). Primer uporabe obstoječega transkripcijskega aparata za lastno transkripcijsko aktivnost je gen *C22orf29*, ki si CpG-promotor in dva 5'-UTR eksona deli z genom *GNB1L* (sl. 14/E). Pri genih brez CpG-otokov znotraj promotorskih regij, je promotorska regija najverjetneje nastala de novo (npr. *RTL1*) (sl. 14/A).

Analizirali smo prisotnost vezavnih motivov za transkripcijske faktorje v promotorskih regijah udomačenih genov (priloga C). Primerjava med različnimi udomačenimi geni ter med človeškimi in mišjimi ortolognimi geni kaže na velike razlike v prisotnosti potencialnih vezavnih mest za transkripcijske faktorje.

### 4.10.3 De novo pridobivanje in evolucija 3' UTR regije

Iz analiz strukture udomačenih genov je razvidna prisotnost na novo pridobljenih 3'neprevedenih (UTR-) regij, ki so jih udomačeni geni pridobili med molekularno udomačitvijo (sl. 12; pregl. 11). Analizirali smo dolžino 3'-UTR regij pri človeških in mišjih referenčnih (RefSeq) genih. Opazili smo velike razlike v dolžini na novo pridobljenih 3'-UTR regij. Najkrajša 3'-UTR regija je pri genu *SCAND3* (281 bp), najdaljša pa pri mišjem genu *PEG10* (5166 bp). Srednja vrednost dolžine 3'-UTR regij genov pri človeku je približno 520 bp (Grillo in sod., 2010). Dolžina 3'-UTR regij pri udomačenih genih je v povprečju precej daljša, saj je srednja vrednost dolžin 3'-UTR udomačenih genov pri človeku približno 1300 bp.

Analiza 3'-UTR zaporedij z orodjem RepeatMasker ni pokazala prisotnosti vrstnospecifičnih TE znotraj teh regij. Nadpovprečno dolge 3'-neprevedene regije pri udomačenih genih torej niso posledica prisotnosti TE in enostavnih ponovitev. Število *»downstream«* odprtih bralnih okvirjev (dORFs) znotraj 3'-UTR zaporedij udomačenih genov je primerljivo s številom dORFs sosednjih medgenskih regij znotraj istega genomskega lokusa in ravno tako ni razlog za povečano velikost 3'-UTR regij.

Geni z občutno podaljšanim 3'-UTR zaporedjem imajo običajno tkivno-specifičen profil izražanja (Stark in sod., 2005). Udomačeni geni bi lahko predstavljali izjemo temu pravilu, saj so udomačeni geni z najdaljšimi 3'-UTR regijami izraženi v večini testiranih tkiv (Unigene EST profil) in jih tako lahko umestimo med hišne (ang. »housekeeping«) gene (pregl. 11).

Table 11: 3' UTR lengths in human and mouse domesticated genes and their expression profiles				
Gen	H. sapiens	Profil izražanja	M. musculus	Profil izražanja
RTL1	>2000	TS	ni podatka	TS
PEG10	5161	HK	5166	TS
LDOC1L	4267	HK	3064	HK
C22ORF29	5029	HK	ni podatka	
ZCCHC5	924	TS	911	TS
ZCCHC16	1584	TS	1912	TS
RGAG4	2268	HK/TS	2521	HK/TS
LDOC1	836	НК	785	TS
FAM127a	821	HK	ni podatka	
FAM127b	835	HK	ni podatka	
FAM127c	1614	HK/TS	ni podatka	
RGAG1	1013	TS	ni podatka	TS
PNMA1	795	HK	735	TS
PNMA2	2981	TS	2835	TS
PNMA3	2023	TS	1969	TS
MOAP1	991	HK	2400	TS
PNMA5	1428	TS	394	TS
PNMA6a	754	TS	ni podatka	
PNMAL1	2070	HK	284	TS
PNMAL2	2367	HK/TS	1750	TS
ZCCHC12	515	TS	518	TS
ZCCHC18	519	TS	1104	TS
CCDC8	865	HK/TS	ni podatka	TS
ARC	1551	TS	1668	TS
GIN1	1898	HK	400	HK/TS
KRBA2	479	TS	ni podatka	
SCAND3	281	TS	ni podatka	
NYNRIN	1842	HK	1776	HK/TS
ASPRV1	568	TS	517	TS

Preglednica 11: Dolžina 3'-UTR regij (bp) pri človeških in mišjih udomačenih genih v povezavi z njihovim profilom izražanja

HK, hišni (ang. »housekeeping«) gen; TS, tkivno-specifičen (ang. »tissue-specific«) gen; Odebeljene črke označujejo spremembo v profilu izražanja med človeškimi in mišjimi ortolognimi geni.

### 4.11 DOLOČANJE VEZAVNIH PARTNERJEV ČLOVEŠKEGA PROTEINA PNMA5 Z DVOHIBRIDNIM SISTEMOM KVASOVKE (Y2H)

Kvasni vektor pGBKT7-PNMA5 z vstavljenim zapisom za človeški protein PNMA5 smo transformirali v sev Y2HGold (Clontech). Vektor pGBKT7-PNMA5 izraža fuzijski protein PNMA5 z DNA vezavno domeno transkripcijskega faktorja *GAL4* (PNMA5-DB). Izražanje zapisa PNMA5-DB v sevu Y2HGold ne sproži avto-aktivacije reporterskih genov in za celice ni toksičeno, kar smo testirali z redčitvenim testom na kontrolnem gojišču ter gojišču z dodanim X- $\alpha$ -Gal in Aureobasidin-om A (AbA).

Sev Y2HGold z vstavljenim vektorjem pGBKT7-PNMA5 smo v tekočem mediju križali s sevom Y187, v katerega je bila predhodno s transformacijo vstavljena knjižnica cDNA pripravljena iz možganov zarodka človeka (Clontech Mate & Plate library – Human fetal brain). Kulturo diploidnih celic smo najprej nanesli na manj strogo selekcijsko gojišče (DDO/-Leu/-Trp/X-α-Gal/AbA). Modre kolonije, ki so zrasle na

manj strogem selekcijskem gojišču, smo nanesli na bolj strogo selekcijsko gojišče (DDO/-Leu/-Trp/-His/-Ade/+X-α-Gal/+AbA). Ker kvasne celice pri transformaciji lahko sprejmejo več različnih vektorjev smo modre kolonije nekajkrat precepili na gojišču DDO/-Leu/-Trp/+X-α-Gal, da bi s segregacijo tarčne vektorje odgovorne za aktivacijo reporterskih genov ločili od morebitnih drugih vektorjev iz cDNA knjižnice.

Da bi identificirali interakcijske partnerje človeškega proteina PNMA5, smo iz kvasnih celic izolirali tarčne vektorje, odgovorne za aktivacijo reporterskih genov, jih namnožili v celicah DH5α ter z sekvenciranjem določili zaporedje njihovih insertov cDNA.

Pridobljena zaporedja cDNA smo analizirali in identificirali 5 zapisov, ki kodirajo potencialne vezavne partnerje PNMA5 (pregl. 12).

Table 12: PNMAS binding partners identified in Y2H experiment					
Oznaka inserta	Opis	Opomba			
SPTBN1	spectrin, beta, non-erythrocytic 1	od 701 do 1207 AK			
SPTBN1	spectrin, beta, non-erythrocytic 1	od 1571 do 2211 AK			
VIM	vimentin				
EPC2	enhancer of polycomb homolog 2				
ATAD3B	ATPase family, AAA-domain containing 3B				

Preglednica 12: Identificirani potencialni vezavni partnerji človeškega proteina PNMA5 Table 12: PNMA5 binding partners identified in Y2H experiment

Kot potencialne vezavne partnerje človeškega proteina PNMA5 smo identificirali proteine SPTBN1, VIM, EPC2 in ATAD3B. Protein PNMA5 je jedrni protein (poglavje 4.12), zato smo na osnovi subcelične lokalizacije kot potencialno lažno pozitivne zadetke označili proteina VIM (intermediarni filament) in ATAD3B (membranski protein, notranja membrana mitohondrijev). Protein EPC2 (histonska acetiltransferaza) smo identificirali kot lažno pozitivne zadetek, saj je izražanje tarčnega vektorja z vstavljenim fragmentom EPC2 v celicah Y2HGold sprožilo izražanje reporterskih genov neodvisno od prisotnosti proteina PNMA5-DB.

Kot vezavna partnerja proteina PNMA5 smo tako identificirali dva različna fragmenta proteina SPTBN1. Prvi fragment SPTBN1 je sestavljen iz spektrinskih ponovitev 5, 6, 7 in 8, drugi pa iz spektrinskih ponovitev 13, 14, 15, 16, 17 in domene PH (sl. 15).

#### A)



#### Slika 15: Shematski prikaz strukture proteina SPTBN1

Z rdečo barvo sta označena dva različna dela proteina SPTBN1 - A) spektrinske ponovitve 5-8. B) spektrinske ponovitve 13-17 in domena PH -, ki smo jih v eksperimentu Y2H identificirali kot vezavna partnerja proteina PNMA5.

Figure 15: Shematic representation of the SPTBN1 protein structure

Red color represents two different parts of the SPTBN1 protein that were identified as a PNMA5 binding proteins in the Y2H experiment. A) spectrin repeats 5-8. B) Spectrin repeats 13-18 and domain PH.

### 4.11.1 Test odvisnosti aktivacije reporterskih genov od prisotnosti PNMA5

Će cDNA fragment iz testirane genomske knjižnice zapisuje (jedrni) protein s sposobnostjo vezave DNA-molekul, lahko AD-domena (aktivacijska domena transkripcijskega faktorja *GAL4*) nastalega fuzijskega proteina aktivira reporterske gene neodvisno od prisotnosti vabe. Za preverjanje odvisnosti aktivacije reporterskih genov od interakcije med proteinom PNMA5 in različnimi identificiranimi fragmenti proteina SPTBN1 smo izvedli redčitveni test. Celice Y2HGold smo ko-transformirali z vektorjem pGBKT7 ali pGBKT7-PNMA5 in z vektorjema, ki nosita zapis za različne dele proteina SPTBN1 (osrednji oziroma C-končni del). Ugotovili smo, da je za aktivacijo reporterskih genov v sevu Y2HGold z različnimi fragmenti SPTBN1 potrebna prisotnost proteina PNMA5 in da različna fragmenta SPTBN1 nista avtoaktivatorja reporterskih genov v sevu Y2HGold.



#### Slika 16: Test odvisnosti aktivacije reporterskih genov od prisotnosti PNMA5

Celice Y2HGold smo transformirali z ustreznimi vektorji in jih gojili v tekočem mediju YNB/-Leu/-Trp. Redčitve (10x) prekonočnih kultur celic smo nanesli na testno in kontrolno gojišče (redčitveni test) ter jih gojili 72 ur. Na kontrolnem gojišču so zrasli vsi testirani sevi (ob prisotnosti PNMA5 in SPTBN1 se kolonije obarvajo modro), na testnem gojišču pa celice brez PNMA5 (prazen vektor pGBKT7) niso zrasle.

#### Figure 16: Bait dependency test

Y2HGold cells were transformed and grown in YNB/-Leu/-Trp media at 10x serial dillutions for 72 hours. Cells containing empty (without PNMA5 insert) pGRBKT7 vector did not grow on the test media.

### 4.11.2 Potrditveni test interakcije med proteinoma PNMA5 in SPTBN1 s koimunoprecipitacijo

Pripravili smo vektor pCDNA3.1 z vstavljenim zapisom za človeški protein PNMA5, ki smo ga na C-koncu označili z V5-značko. Vektor pCDNA3.1-PNMA5-V5 smo izrazili v celicah SHSY-5Y. Po 72 urah smo pripravili lizate transfeciranih celic in ne-transfeciranih celic ter celic, ki so bile 24 ur inkubirane v mediju z dodanim rekombinantnim TGF $\beta$ 1 (25 ng/ml). Izolirano jedrno frakcijo lizata celic smo uporabili za ko-imunoprecipitacijo. Lizate celic smo inkubirali z mišjim protitelesom proti V5 (Invitrogen) in nato dodali magnetne kroglice z vezanim proteinom G (Active Motif). Frakcijo proteinov, vezanih na magnetne kroglice, smo po ko-imunoprecipitaciji nanesli na NaDS-PAGE in izvedli prenos western. Po prenosu western nam na nitrocelulozni membrani z zajčjim protitelesom proti SPTBN1 (Santa Cruz Biotechnology) ni uspelo

zaznali prisotnosti proteina SPTBN1, tako da nam interakcije med človeškim PNMA5-V5 in nativnim SPTBN1 v celicah SHSY-5Y ni uspelo potrditi.

### 4.12 SUB-CELIČNA LOKALIZACIJA PROTEINA PNMA5

Zapis za človeški protein PNMA5 smo vstavili v ekspresijski vektor pCDNA3, ki v sesalskih celicah omogoča izražanje proteina, na C-koncu označenega z EGFP. Pripravljeni vektor smo vstavili v celice HEK293T in opazovali lokalizacijo proteina PNMA5-EGFP pod fluorescenčnim mikroskopom. Jedra celic HEK293T smo obarvali z barvilom DAPI. Rekombinantni protein PNMA5-EGFP se nahaja v jedru celic HEK293T (sl. 17).



Slika 17: Sub-celična lokalizacija človeškega proteina PNMA5-EGFP v celicah HEK293T

Fotografija je bila zajeta na fluorescenčnem mikroskopu pri 100x povečavi. Jedra celic HEK293T so označena z barvilom DAPI (Invitrogen).

**Figure 17: PNMA5 subcellular localization in HEK293T cells** HEK293T cells stained with DAPI (Invitrogen) at 100x magnification.

Protein PNMA5-EGFP se v jedru celic HEK293T nahaja v obliki številnih zrnatih struktur, ki jih je od približno 5 do približno 20 na celico/jedro. Vzorec sub-celične lokalizacije proteina PNMA5-EGFP ustreza vzorcu lokalizacije nekaterih znanih jedrnih struktur, med katerimi so tudi PML-jedrna telesca.





#### Slika 18: Protein PNMA5-EGFP je ko-lokaliziran z jedrnimi telesci PML

A) Slike preparata celic HEK293T, ki izražajo rekombinantni človeški protein PNMA5-EGFP in so označene z mišjimi protitelesi proti PML smo posneli s fluorescenčnim mikroskopom pri 100x povečavi. Za označitev jeder smo uporabili barvilo DAPI. Razvidni sta jedrna lokalizacija človeškega proteina PNMA5-EGFP in ko-lokalizacija proteina PNMA5-EGFP z jedrnimi telesci PML. Prekrivajoče se strukture PNMA5-EGFP in PML-telesc so na združeni sliki rumene barve. B) Slike preparata celic SHSY-5Y, ki izražajo rekombinantni človeški protein PNMA5-EGFP in so označene z mišjimi protitelesi proti PML smo posneli s konfokalnim mikroskopom pri 68x povečavi (Z-projekcija, maksimalna intenziteta). Razvidna je delna ko-lokalizacija proteina PNMA5-EGFP in jedrnih telesc PML. V jedru SHSY-5Y celic opazimo posamezne proste (ne ko-lokalizirane) strukture PNMA5 in PML.

#### Figure 18: PNMA5-EGFP colocalizes with PML nuclear bodies

A) HEK293T cells expressing recombinant human PNMA5-EGFP protein stained with mouse anti-PML antibodies at 100x magnification. Nuclei were stained with DAPI. B) SHSY-5Y cells expressing recombinant human PNMA5-EGFP protein stained with mouse anti-PML antibodies at 63x magnification (Z-projection, max intensity).

Preparat fiksiranih celic HEK293T, ki izražajo protein PNMA5-EGFP, smo inkubirali skupaj s protitelesi proti anti-PML, ter opazovali lokalizacijo proteina PNMA5-EGFP in označenih PML-struktur pod fluorescenčnim mikroskopom (sl. 18/A). Razvidno je, da je človeški protein PNMA5-EGFP ko-lokaliziran s PML-jedrnimi telesci v jedru celic HEK293T.

Protein PNMA5-EGFP se v celicah SHSY-5Y lokalizira v jedru (sl. 18/B). Imunodetekcija PML-jedrnih telesc pri celicah SHSY-5Y kaže na delno ko-lokalizacijo struktur PNMA5-GFP in PML-jedrnih telesc. V jedru transfeciranih celic SHSY-5Y opazimo tako ko-lokalizirana PML-jedrna telesca in strukture PNMA5-EGFP kot prosta PML jedrna telesca in proste strukture PNMA5-EGFP.

### 4.13 SUB-CELIČNA LOKALIZACIJA PROTEINA SPTBN1

Endogeni človeški protein SPTBN1 smo zaznali z zajčjimi protitelesi proti SPTBN1 (Santa Cruz Biotechnology). V celicah SHSY-5Y protein SPTBN1 zaznamo predvsem v citoplazmi celic, v manjši meri pa tudi v jedru (sl. 19/A). Inkubacija celic SHSY-5Y z rekombinantnim TGF $\beta$ 1 (25 ng/ml, 24 ur) povzroči delno translokacijo proteina SPTBN1 iz citoplazme na plazemsko in jedrno membrano (sl. 19/B).



B)



#### Slika 19: Sub-celična lokalizacija proteina SPTBN1

Celice SHSY-5Y smo 48 ur gojili v mediju brez dodanega TGF $\beta$ 1 (A) oziroma v mediju z dodanim TGF $\beta$ 1 (25 ng/ml) (B). Preparate celic SHSY-5Y smo fiksirali z metanolom ter izvedli imunodetekcijo s primarnimi protitelesi proti SPTBN1 in sekundarnimi protitelesi Alexa Fluor 488. Slike smo posneli s konfokalnim mikroskopom pri 68x povečavi.

#### Figure 19: SPTBN1 subcellular localization

SHSY-5Y cells were grown in media without TGF  $\beta$ 1 (A) or with TGF $\beta$ 1 (25 ng/ml) (B) for 48 hours. Primary anti-SPTBN1 antibodies were used, followed by staining with Alexa Fluor 488 secondary antibodies.

### 4.14 AKTIVNOST SMAD SIGNALNE POTI V ODVISNOSTI OD IZRAŽANJA PROTEINA PNMA5

Citokin TGF $\beta$ 1 vpliva na SMAD-celično signaliziranje prek interakcije z membranskimi receptorji za TGF $\beta$ , kar vodi do fosforilacije in s tem aktivacije citoplazemskih proteinov SMAD, ki delujejo kot transkripcijski faktorji (Massagué, 2012). Morebitni vpliv proteina PNMA5 na aktivnost SMAD-signalne poti smo testirali s čezmernim izražanjem proteina PNMA5 v celicah HEK293T.

Sesalski ekspresijski vektor pcDNA3-PNMA5, ki izraža človeški protein PNMA5, smo v celice HEK293T vstavili skupaj z reporterskim vektorjem p3TP-Lux (SMAD-odzivni promotor) in kontrolnim vektorjem pRL-SV40 (konstitutivni promotor). Celice smo 24 ur po transfekciji tretirali s TFGβ1 (10 ng/ml, 24 ur) in pripravili lizate celic. Izmerili smo aktivnost luciferaze kresnice (ang. »Firefly«) in luciferaze organizma *Renilla reniformis*, pri čemer smo aktivnost luciferaze Renilla uporabili kot interno kontrolo za normalizacijo rezultatov.



Slika 20: Aktivnost SMAD signalne poti v odvisnosti od izražanja proteina PNMA5 v celicah HEK293T

Prikazana je sprememba aktivnosti luciferaze iz kresnice izražene iz reporterskega vektorja p3TP-Lux. Eksperiment je bil izveden v treh bioloških ponovitvah (srednja vrednosti  $\pm$  standardna deviacija).

#### Figure 20: PNMA5 dependant SMAD signaling upregulation in HEK293T cells

Activity of Firefly luciferase expressed from p3TP-Lux plasmid (fold change) is shown (3 biological replicates (mean  $\pm$  standard deviation)).

Ugotovili smo, da tako tretiranje celic s TGFβ1 kot čezmerno izražanje proteina PNMA5 v celicah HEK293T poveča aktivnost SMAD-signalne poti (sl. 20). Luciferazna aktivnost je pri celicah, ki jim v medij dodamo TGFβ1 in hkrati izražajo protein PNMA5, v primerjavi s kontrolnim vzorcem 2-krat višja. Sklepamo lahko, da imata izražanje proteina PNMA5 in delovanje TGFβ1 skupen pozitiven sinergistični vpliv na povišano aktivnost SMAD-signalne poti.

# 5 RAZPRAVA

Dostopnost celotnih genomov sesalcev in mnogih drugih strunarjev je omogočila filogenomsko analizo genov in multigenskih družin, ki so nastali iz ostankov retroelementov. S filogenomsko analizo smo pridobili podatke o številu in razširienosti udomačenih genov, njihovi strukturi, kromosomski lokaciji ter organizaciji njihovih kromosomskih lokusov. Pri iskanju novih genov, nastalih iz različnih domen retrotranspozonov (gag, integraza, proteaza), smo uporabili prosto dostopne genomske baze podatkov. Prisotnost udomačenih genov smo analizirali pri različnih skupinah placentalnih sesalcev (Boreoeutheria, Afrotheria in Xenarthra), ključnih vretenčarjih (dvoživke, plazilci, ptiči) in drugih strunarjih. Preiskali smo več kot 90 različnih genomov strunarjev iz različnih taksonomskih skupin, kar je bilo ključnega pomena za določitev izvora in časa nastanka (udomačitve) novonastalih genov. Zanimalo nas je, ali imajo družine udomačenih genov (predvsem Sushi in PNMA) enoten izvor – ali so monofiletskega oziroma parafiletskega izvora, oziroma koliko molekularnih udomačitev je bilo potrebnih za njihov nastanek. Zanimal nas je mehanizem nastanka udomačenih genov, iz katerih genomskih elementov so nastali in kako so pridobivali regulatorna zaporedja, ki so omogočila neofunkcionalizacijo novonastalih genov.

Odprti bralni okvir brez stop-kodonov in mutacij, ki povzročijo premik bralnega okvira, je dober pokazatelj evolucijske in funkcionalne ohranjenosti gena. S primerjavo nukleotidnih zaporedij udomačenih genov smo lahko razlikovali med potencialnimi funkcionalnimi geni in zaporedji potencialnih psevdogenov. V nekaterih primerih smo identificirali fragmentirana in nepopolna nukleotidna in proteinska zaporedja udomačenih genov –predvsem pri organizmih s slabšo pokritostjo genomskih podatkov. V teh primerih razlikovanje med prisotnostjo funkcionalnega gena in okvarjenimi zaporedji (psevdogeni) ni bilo zanesljivo.

# 5.1 ANALIZA RAZŠIRJENOSTI UDOMAČENIH GENOV

Uporaba širokega taksonomskega vzorčenja nam je omogočila pridobiti obsežne podatke o udomačenih genih v genomih različnih skupin sesalcev, vretenčarjev in strunarjev (priloga A). Predhodne študije udomačenih genov so bile omejene samo na analizo prisotnosti udomačenih genov v genomih placentalnih sesalcev iz skupine Boreoeutheria (Llorens in Marín, 2001; Lynch in Tristem, 2003; Gorinšek in sod., 2004; Brandt in sod., 2005a; Brandt in sod., 2005b; Zdobnov in sod., 2005; Campillos in sod., 2006), nič pa ni bilo znanega o prisotnosti teh genov znotraj skupin placentalnih sesalcev Xenarthra in Afrotheria, zato tudi sklepanje o času nastanka teh genov ni bilo mogoče. Z analizo več genomov iz različnih skupin placentalnih sesalcev smo dokazali, da je večina analiziranih udomačenih genov, nastalih iz skupine Ty3/Gypsy retrotranspozonov, prisotna v vseh treh skupinah placentalnih sesalcev (Afrotheria,

Xenarthra in Boreoeutheria). Dokazali smo, da so v genomih placentalnih sesalcev široko razširjeni geni iz družin Sushi in PNMA, geni, nastali iz integrazne domene retrotranspozonov ter gena *ARC* in *ASPVR1*. V genomih bazalnih sesalcev (Monotremata in Marsupialia) ter v genomih plazilcev, ptičev in dvoživk so udomačeni geni bistveno manj razširjeni. Omejeno razširjenost udomačenih genov smo opazili tudi v genomih bazalnih vretenčarjev (obloustke, hrustančnice in ribe kostnice), kjer sta prisotna samo dva gena, ki sta nastala iz integrazne domene (*GIN1* in *GIN2*).

Predhodne študije zaradi omejenega taksonomskega vzorčenja (Llorens in Marín, 2001; Lynch in Tristem, 2003; Gorinšek in sod., 2004; Brandt in sod., 2005a; Brandt in sod., 2005b Zdobnov in sod., 2005; Campillos in sod., 2006) niso podale celovitega odgovora o času nastanka različnih skupin udomačenih genov. Ob uporabi širokega taksonomskega vzorčenja smo lahko iz razširjenosti udomačenih genov znotraj različnih taksonomskih skupin sklepali na čas in dinamiko nastanka (družin) udomačenih genov. Iz podatkov o razširjenosti udomačenih genov smo dokazali, da je do nastanka, širitve in diverzifikacije večine iz gag domene nastalih udomačenih genov prišlo šele v predniku placentalnih sesalcev pred približno 176 milijoni let. Geni, ki so nastali iz integrazne in proteazne domene retrotranspozonov, so evolucijsko starejši (*GIN1* in *GIN2* sta nastala v predniku strunarjev pred več kot 500 milijoni let, *ASPRV1* je nastal v predniku sesalcev pred približno 220 do 250 milijoni let), vendar pa so v primerjavi z udomačenimi geni nastalimi iz domene gag prisotni v bistveno manjšem številu in so bistveno manj diverzificirani.

# 5.2 FILOGENOMSKA ANALIZA UDOMAČENIH GENOV

Za pojasnitev ortolognih in paralognih odnosov med udomačenimi geni smo uporabili metode filogenetske analize in analizo ohranjenih kromosomskih položajev (analiza sinteničnih kromosomskih položajev). Razlikovanje med paralogi in ortologi nam je podalo vpogled v število različnih udomačenih genov pri posameznih organizmih in v evolucijo multigenskih družin udomačenih genov.

Za pripravo filogenetskih dreves smo poravnali reprezentativen izbor zaporedij udomačenih genov prisotnih iz organizmih različnih taksonomskih skupin. Zaradi nizke podobnosti ortolognih genov znotraj posameznih multigenskih družin (Sushi in PNMA) smo poravnali samo najbolj ohranjene dele zaporedij (npr. domeno gag). Ugotovili smo, da uporaba metod NJ in ML pri filogenetski analizi udomačenih genov poda primerljive filograme s podobno visoko ločljivostjo podatkov, vendar pa z relativno nizkimi vrednostmi zanesljivosti vejitev. Z uporabo Bayesove metode filogenetske analize smo zgradili filograme s primerljivo visoko ločljivostjo podatkov, vendar z višjimi vrednostmi zanesljivosti vejitev kot pri uporabi metod NJ in ML. Filogenetska analiza je izboljšala dosedanjo sliko evolucijskih odnosov znotraj družin udomačenih genov ter nam podala nov in izboljšan vpogled v evolucijo teh genov. V predhodnih študijah je bilo zaradi nedostopnosti genomskih podatkov analizirano le manjše število zaporedij udomačenih genov iz ene same skupine placentalnih sesalcev (Boreoeutheria) (Brandt in sod., 2005a; Brandt in sod., 2005b; Campillos in sod., 2006), zaradi česar ni bilo mogoče določiti izvora udomačenih genov, dinamike njihovega nastajanja ter odgovoriti na vprašanje o mono- ali para-filetskem izvoru družine Sushi (Brandt in sod., 2005b; Volff, 2006).

Izvor udomačenih genov v predhodnih študijah ni bil natančno določen (Llorens in Marín, 2001; Brandt in sod., 2005a; Brandt in sod., 2005b; Campillos in sod., 2006; Volff, 2006), zato smo s filogenetsko analizo določili izvorne elemente proučevanih udomačenih genov. Potrdili smo, da je družina udomačenih genov Sushi nastala iz tipa Ty3/Gypsy LTR-retrotranspozonov z vključeno kromo-domeno (Kromovirusi) (Gorinšek in sod., 2004; Brandt in sod., 2005a; Kordiš in sod., 2005). Ugotovili smo, da so geni iz družine PNMA nastali iz elementov klade Barthez, gen *ARC* iz elementov klade Osvaldo, gen *ASPRV1* iz elementov klade Cigr2, integrazni geni pa so nastali iz elementov klade Gmr1, oziroma pri genu *NYNRIN* iz integrazne domene endogenih retrovirusov (ERV) (Marco in Marín, 2009). S filogenetsko analizo smo tako dokazali, da so družine udomačenih genov nastale v več neodvisnih dogodkih molekularne udomačitve ter so se naprej širile in diverzificirale z genskimi podvojitvami.

Na podlagi filogenetske analize in strukturne analize genov družine Sushi sklepamo, da je družina udomačenih genov Sushi nastala kot posledica vsaj dveh oziroma treh neodvisnih dogodkov molekularne udomačitve. Gena *PEG10* in *RTL1* sta najverjetneje nastala iz izvornih retroelementov polne velikosti, saj poleg domene gag ohranjata tudi proteazno domeno (*PEG10*) oziroma proteazno domeno in reverzno transkriptazo (*RTL1*). Ostali geni iz družine Sushi imajo (delno) ohranjeno samo domeno gag in so najverjetneje nastali iz ostankov domene gag Kromovirusov (sl. 10/C/D).

V primeru družine PNMA na podlagi filogenetske analize, strukturne analize genov in analize ohranjene sintenije sklepamo, da je genska družina PNMA monofiletskega izvora. Iz izvornega gena *M-PNMA*, ki je prisoten pri vrečarjih, je v predniku placentalnih sesalcev z genskimi podvojitvami najverjetneje nastalo več kopij izvornega gena, iz katerih so nastali geni genske družine PNMA. Na podlagi strukturnih značilnosti in kromosomskih položajev lahko gensko družino PNMA razdelimo na tri podskupine genov: a) *ZCCHC12/ZCCHC18*; b) *PNMAL1, PNMAL2, CCDC8*; c) *PNMA1, MOAP1, PNMA2, PNMA3, PNMA5* in *PNMA6a/b* (sl. 10/A/B).

Vzorec nastanka udomačenih genov z integrazno domeno se razlikuje od vzorca nastanka udomačenih genov, nastalih iz domene gag. Udomačena gena *GIN1* in *GIN2*, katerih integrazna domena je podobna integrazi elementov GINGER (Marín, 2010), sta

nastala pred več kot 500 milijoni let in sta tako bistveno starejša od večine udomačenih genov, nastalih iz domene gag retrotranspozonov. DNA-transpozicijski elementi skupine GINGER, iz katerih sta nastala GIN1 in GIN2, naj bi integrazno domeno pridobili od skupine retrotranspozonov Ty3/Gypsy (Marín, 2010). Gena KRBA2 in SCAND3 naj bi nastala v procesu molekularne udomačitve elementov GINGER2, ki so bili s horizontalnim prenosom preneseni iz žuželk redu Hemiptera (stenice) v prednika sesalcev (Llorens in sod., 2012). Rezultati naše filogenetske analize potrjujejo podobnost elementov GINGER z integraznimi geni kopenskih vretenčarjev, vendar hkrati kažejo podobnost integrazne domene teh genov z integrazno domeno elementov Metaviridae (sl. 8). Znano je, da se domena SCAN nahaja v regiji gag retroelementov klade Gmr1 iz skupine Ty3/Gypsy, prisotnih pri Deuterostomia (Goodwin in Poulter, 2002; Kordiš, 2009; Emerson in Thomas, 2011). Glede na strukturno organizacijo evolucijsko starih elementov Gmr1, ki vključujejo tako SCAN- kot integrazno domeno, sklepamo, da sta gena KRBA2 in SCAND3 nastala iz ostankov elementov Gmr1 v predniku Theria. Podobnost integrazne domene nekaterih udomačenih genov z integrazno domeno elementov GINGER2 je po našem mnenju posledica konvergentne evolucije integraznih domen genov KRBA2 in SCAND3 ter integraznih domen elementov GINGER2 pri žuželkah, kar se zdi verjetnejši scenarij kot horizontalni prenos elementov GINGER2 iz žuželk v prednika sesalcev.

# 5.3 IZJEMNA OHRANJENOST ZAPOREDIJ IN KROMOSOMSKIH POLOŽAJEV UDOMAČENIH GENOV

Primerjali smo ohranjenost ortolognih in paralognih zaporedij udomačenih genov pri sesalcih in drugih vretenčarjih. Primerjava ohranjenosti med ortolognimi zaporedji udomačenih genov kaže na visoko stopnjo ohranjenosti (pregl. 7). Ohranjenost ortolognih zaporedij PNMA in Sushi med različnimi vrstami placentalnih sesalcev je višja kot ohranjenost paralognih zapisov znotraj posamezne vrste. Nizka ohranjenost med paralognimi zaporedji znotraj placentalnih sesalcev ob hkratni visoki ohranjenosti med ortolognimi zaporedji nakazuje na hitro adaptivno evolucijo teh zaporedij v predniku placentalnih sesalcev.

Poleg ohranjenosti zaporedij udomačenih genov smo raziskali tudi ohranjenost njihovih kromosomskih položajev. Analiza ohranjenosti kromosomskih lokusov (analiza sintenije) udomačenih genov pri različnih organizmih nam je dala nedvoumen odgovor o paralognih/ortolognih odnosih med posameznimi udomačenimi geni znotraj multigenskih družin ter pokazala, da so bili tako udomačeni geni kot njihovi kromosomski položaji ohranjeni že v predniku placentalnih sesalcev. Dobro pojasnjeni evolucijski odnosi znotraj multigenskih družin udomačenih genov in analiza njihovih ohranjenih kromosomskih položajev omogočajo analizo kromosomske mobilnosti udomačenih genov v času nastajanja teh multigenskih družin.

Združena analiza filogenetske razširjenosti, časa in dinamike nastanka udomačenih genov ter ohranjenosti teh zaporedij in njihovih kromosomskih položajev nakazuje na hitro širitev multigenskih družin udomačenih genov, nastalih iz domene gag v predniku placentalnih sesalcev in na hitro adaptivno evolucijo teh genov med neofunkcionalizacijo. Opažanja so skladna tudi z dejstvom, da so bili udomačeni geni družine Sushi zaradi na novo pridobljene funkcionalne vloge podvrženi močni negativni selekciji med evolucijo placentalnih sesalcev (Brandt in sod., 2005b, Volff, 2006).

## 5.4 KROMOSOMSKA MOBILNOST UDOMAČENIH GENOV

Znano je, da je velik delež udomačenih genov – predvsem iz družine Sushi in polovica genov iz družine PNMA pri človeku na spolnem kromosomu X (Brandt in sod., 2005a; Brandt in sod., 2005b; Campillos in sod., 2006). Dinamika kromosomske mobilnosti udomačenih genov v preteklosti ni bila podrobno raziskana. Podatki o izvornih genih multigenskih družin udomačenih genov, predvsem podatki o udomačenih genih in njihovih kromosomskih položajih v genomih vrečarjev in placentalnih sesalcev iz razredov Xenarthra in Afrotheria, so nam dali poglobljen vpogled v dinamiko njihovega nastajanja in njihovo kromosomsko mobilnost.

Analiza kromosomske mobilnosti je pokazala na visoko dinamiko nastajanja in premikanja udomačenih genov v predniku placentalnih sesalcev, saj smo v primeru genske družine PNMA zaznali številne primere premikov udomačenih genov z avtosomov na kromosom X in z kromosoma X na avtosomalne kromosome (poglavje 4.7). Po hitri fiksaciji na novo nastalih udomačenih genov v predniku placentalnih sesalcev so njihovi kromosomski položaji znotraj placentalnih sesalcev ostali močno ohranjeni, kar je razvidno iz analize ohranjenih kromosomskih položajev.

Prisotnost velikega dela udomačenih genov iz družin Sushi in PNMA na spolnem kromosomu X nakazuje na njihovo morebitno vlogo v razvoju s spolom povezanih funkcij. Za kromosom X je znano, da vsebuje številne gene, ki se izražajo v spolnih celicah in imajo s spolom povezane funkcije (Wang in sod., 2001; Khil in sod., 2004; Parsch in Ellegren, 2013). Izražanje novonastalih genov je pogosto povečano v testisih (Kaessmann, 2010). V primeru novih retrogenov je bilo dokazano, da kromosom X preferenčno akumulira gene, ki so udeleženi v razvoj spolnih funkcij in reprodukcijo (Potrzebowski in sod., 2010; Kaessmann, 2010).

## 5.5 FILOGENETSKA ANALIZA PROGENITORJEV UDOMAČENIH GENOV (ELEMENTOV METAVIRIDAE)

Za razumevanje evolucijskih mehanizmov, dinamike in časa nastanka udomačenih genov, je ključnega pomena dobro poznavanje dinamike in razširjenosti njihovih izvornih elementov. S Filogenetsko analizo retroelementov iz skupine Metaviridae pri

Deuterostomia smo lahko določili, katere družine oziroma tipi retroelementov so bili izvorni elementi pri nastanku posameznih genov in multigenskih družin udomačenih genov.

Filogenetska analiza elementov Metaviridae pri Deuterostomia je podala izboljšano sliko razširjenosti teh elementov znotraj različnih skupin kopenskih vretenčarjev. Dokazali smo, da so v genomih različnih plazilcev prisotni številni aktivni elementi iz različnih skupin družine Metaviridae (poglavje 4.8). Prejšnje študije elementov Metaviridae pri Deuterostomia so bile omejene z nedostopnostjo genomov ključnih plazilcev (Sauropsida) (Gorinšek in sod., 2004; Kordiš, 2005; Kordiš, 2009).

Čeprav elementov Metaviridae v genomih sodobnih sesalcev ni (Lander in sod., 2001; Volff in sod., 2003; Gorinšek in sod., 2004; Kordiš, 2005; Böhne in sod., 2008; Kordiš, 2009), je iz filogenetske analize razvidno, da so bili aktivni elementi Metaviridae prisotni v genomih njihovih prednikov. Odsotnost izvornih transpozicijskih elementov v genomih sesalcev in dolgo obdobje, ko so bili elementi Metaviridae prisotni v prednikih sesalcev samo v obliki okvarjenih (fragmentiranih), neaktivnih zaporedij, je ključnega pomena za razumevanje procesa nastanka udomačenih genov (procesa molekularne udomačitve) in ima pomembne posledice glede hipotez o omejevanju širjenja starševskih transpozicijskih elementov v genomih gostitelja s strani udomačenih genov (Llorens in Marin, 2001; Lynch in Tristem, 2003; Brandt in sod., 2005a; Brandt in sod., 2005b; Campillos in sod., 2006).

Hipoteze, da lahko nekateri udomačeni geni, npr. *GIN1* in *RTL1*, omejijo kolonizacijo gostiteljskih genomov s strani izvornih transpozicijskih elementov ter povzročijo njihovo izumrtje, so se izkazale za napačne, saj so udomačeni geni, ki so nastali iz integrazne domene in domene gag retrotranspozonov pri sesalcih, nastali iz ostankov neaktivnih retrotranspozonov. Nabor transpozicijskih elementov v predniku sodobnih sesalcev je bil – v primerjavi s prednikom Synapsida, ki je imel podobno bogat nabor TE kot predniki Sauropsida – že močno spremenjen kot posledica velikega izumrtja na prehodu med permom in triasom (pred približno 250 milijoni let) (Kordiš in sod., 2006; Kordiš, 2009).

Primerjava analiz razširjenosti udomačenih genov in razširjenosti njihovih izvornih elementov je pokazala, da je večina analiziranih udomačenih genov nastala iz ostankov neaktivnih retroelementov v predniku placentalnih sesalcev. Aktivnih elementov Metaviridae v genomih sesalcev ni, zato tudi niso mogli biti neposreden vir zaporedij za nastanek novih udomačenih genov, kot je bilo pogosto napačno predlagano (npr. izvor gena *SIRH12* pri vrečarjih (Ono in sod., 2012)).

# 5.6 DE NOVO PRIDOBIVANJE PROMOTORSKIH IN DRUGIH REGULATORNIH ZAPOREDIJ

Mehanizem ter izvor promotorskih zaporedij in drugih regulatornih regij udomačenih genov, nastalih iz retroelementov, v dosedanjih študijah še ni bil pojasnjen (Brandt in sod., 2005a; Brandt in sod., 2005b; Schüller in sod., 2005; Campillos in sod., 2006), oziroma je bil pojasnjen samo v primeru nekaterih udomačenih genov, katerih transkripcijo omogoča prisotnost dvo-smernega promotorja v njihovi bližini (Kalitsis in Safferty, 2009).

Na osnovi primerjave razširjenosti elementov Metaviridae in časa nastanka udomačenih genov sklepamo, da so ti geni nastali iz ostankov neaktivnih retroelementov, ki so bili v genomu gostiteljev prisotni daljši čas (približno 100 milijonov let). Ker ostanki neaktivnih retroelementov nimajo regulatornih zaporedij (Kaessmann in sod., 2009; Kaessmann, 2010), so ta zaporedja, da bi postala neofunkcionalizirana, morala na novo pridobiti promotorske in druge regulatorne regije. Preživetje in ohranitev teh zaporedij v genomih gostiteljev sta bila pogojena s pridobivanjem regulatornih zaporedij ter morebitnih drugih selekcijskih pritiskov, zaradi katerih zaporedja niso postala psevdogenizirana in se niso izgubila iz genomov.

Proces molekularne udomačitve oziroma eksaptacije ostankov neaktivnih transpozicijskih elementov je vključeval različne mehanizme, ki so vodili do eksonizacije teh zaporedij in pridobivanja novih regulatornih regij (poglavje 4.10). Pomembno vlogo pri regulatorni evoluciji udomačenih genov sta igrala pridobivanje in evolucija 5'- in 3'-neprevedenih zaporedij teh genov. Na evolucijo novih neprevedenih zaporedij je pomembno vplival tudi proces pridobivanja novih intronov znotraj teh regij, saj so s tem podaljšanjem udomačeni geni pridobili dostop do oddaljenih (funkcionalnih) promotorskih zaporedij (Kordiš, 2011; Kordiš in Kokošar, 2012).

Analiza transkripcijske aktivnosti udomačenih genov je pokazala, da imajo številni udomačeni geni povečano izražanje v testisih (Brandt in sod., 2005a, Takaji in sod., 2009, NCBI Unigene), kar je najverjetneje povezano z odprtim stanjem kromatina v tem tkivu (Kaessmann, 2010). Analiza kromosomskih položajev in ohranjenih sintenij udomačenih genov je pokazala, da se je fiksacija ostankov retroelementov in s tem nastanek udomačenih genov najpogosteje dogajal v genomskih regijah z višjim transkripcijskim potencialom (CpG-bogate regije) (pregl. 10). Pogosto so udomačeni geni pridobili promotorje, bogate s CpG dinukleotidi, ki niso bili del promotorske regije obstoječih genov. Evolucija intronsko/eksonske-strukture 5'-neprevedenih zaporedij je udomačenim genom omogočila pridobitev promotorskih zaporedij, ki so v nekaterih primerih lahko zelo oddaljena (človeški gen *ZCCHC16* ima znotraj 5'-UTR zaporedja dva introna, dolga približno 48 kb in 323 kb). Poleg pridobitve novih promotorskih zaporedij je izražanje udomačenih genov lahko pogojeno z uporabo obstoječih dvo-

smernih promotorjev (Kalitsis in Safferty, 2009), oziroma skupnega promotorja in 5'-UTR eksonov (gen *C22orf29*) (sl. 14/e).

Opažena raznolikost v prisotnosti vezavnih motivov za transkripcijske faktorje v promotorskih regijah udomačenih genov pri miši in človeku (priloga C) se ujema z velikimi razlikami v profilih izražanja udomačenih genov (tako znotraj vrste kot med različnimi vrstami) (Brandt in sod., 2005a; Takaji, 2009; NCBI Unigene) ter nakazuje, da so promotorske/regulatorne regije udomačenih genov podvržene močni *cis*-regulatorni evoluciji.

Analiza strukture in dolžine 3'-neprevedenih zaporedij udomačenih genov je pokazala, da se 3'-UTR regije udomačenih genov zelo razlikujejo po velikosti (pregl. 11), velik del teh genov pa ima 3'-UTR zaporedja, ki so bistveno daljša od tipične dolžine pri človeku (približno 520 bp) (Grillo in sod., 2010). 3'-UTR zaporedja imajo pomembno vlogo v regulaciji izražanja genov, saj prek različnih molekularnih procesov (stabilnost transkripta, učinkovitost translacije, poliadenilacija, subcelična lokalizacija in transport mRNA) uravnavajo pravilno časovno in prostorsko izražanje genov (Andreassi in Riccio, 2009; Barett in sod., 2012). Pridobivanje 3'-UTR zaporedij udomačenih genov je verjetno imelo pomembne posledice za evolucijo uravnavanja izražanja teh genov, pridobivanje njihovih vrstno, tkivno in časovno specifičnih vzorcev izražanja (Brandt in sod., 2005a, Takaji in sod., 2009, NCBI Unigene) ter njihovo vključitev v obstoječa genska regulatorna omrežja.

## 5.7 PRIDOBIVANJE NOVIH FENOTIPSKIH VLOG UDOMAČENIH GENOV

Filogenomska analiza nastanka, razširjenosti in evolucije multigenskih družin udomačenih genov je pokazala, da je večina udomačenih genov nastala v predniku placentalnih sesalcev. Pojav oziroma nastanek udomačenih genov časovno sovpada z razvojem fenotipskih novosti pri placentalnih sesalcih, kot sta placenta (Kaneko-Ishino in Ishino, 2012) in neokorteks (Oldham in sod., 2006, Takaji in sod., 2009). Podatki iz literature kažejo, da so nekateri udomačeni geni ključno sooblikovali, oziroma celo omogočili razvoj pomembnih fenotipskih novosti njihovih gostiteljev (Kaneko-Ishino in Ishino, 2012). Iz tega vidika sta še posebno zanimiva gena *PEG10* in *RTL1*, ki imata ključno vlogo v razvoju placente in stika med placento in zarodkom, ter tako pomembno vplivata na normalen embrionalni razvoj zarodka (Kaneko-Ishino in Ishino, 2012). Gen *ARC*, ki je nastal že v predniku kopenskih vretenčarjev pred približno 395 milijoni let, je drugi ključni primer udomačitve retrotranspozicijskega zaporedja, saj ima pomembno vlogo pri ohranjanju plastičnosti sinaps in je pri miših nujno potreben za tvorbo dolgoročnega spomina (Lyford in sod., 2005; Plath in sod., 2006; Korb in Finkbeiner, 2011). Izražanje genov *PEG10* in *RTL1* je pri placentalnih sesalcih monoalelno in uravnavano prek diferencialne metilacije očetovega oziroma materinega alela z mehanizmom vtisnjenja (ang. »imprinting«). Gen *PEG10* je bil prvi vtisnjeni gen, za katerega je bilo dokazano, da je njegov mehanizem vtisnjenja skupen placentalnim sesalcem (Ono in sod., 2001) in vrečarjem (Suzuki in sod., 2007). Prej so namreč domnevali, da mehanizem vtisnjenja pri placentalnih sesalcih in vrečarjih nima skupnega izvora (Suzuki in sod, 2005). Metilacija genomske regije *PEG10* in posledično evolucija vtisnjenja tega gena pri Theria naj bi bili posledica epigenetskega utišanja izvornega retrotranspozona, vstavljenega v ta genomski lokus (Suzuki in sod., 2007). Epigenetsko utišanje z metilacijo je pogost mehanizem obrambe organizmov pred aktivnostjo tuje oziroma parazitske DNA (Barlow, 1993; Yoder in sod., 1997; Walsh in sod., 1997), vendar se postavlja vprašanje, ali je bila metilacija izvornega retroelementa iz katerega je gen *PEG10* nastal res odločujoči dejavnik za razvoj vtisnjenja te genomske regije, saj v času nastanka gena *PEG10* zaporedje njegovega izvornega retroelementa že dolgo ni bilo več aktivno.

Za nekatere udomačene gene je značilno, da so pogosto izraženi pri različnih patoloških stanjih, predvsem pri rakastih obolenjih. Skladno s tem je bila prepoznana biološka vloga nekaterih udomačenih genov iz družin Sushi (Okabe in sod., 2003; Nagasaki in sod., 2003; Inoue in sod., 2005) in PNMA (Fu in sod., 2007; Chen in D'Mello, 2010; Dai in sod., 2011) pri regulaciji apoptoze. Strukturne značilnosti proteinov, ki jih kodirajo geni iz družin Sushi in PNMA, so si lahko podobne, saj vsi izhajajo iz domene gag retrotranspozonov. Ker pa ima ta domena velik interakcijskih potencial (poglavje 2.9), je bila v proteinih družin Sushi in PNMA lahko uporabljena kot interakcijski modul za vezavo na proteine, ki so udeleženi v podobnih celičnih procesih – v tem primeru na proteine, udeležene pri regulaciji apoptoze. Vloga nekaterih proteinov družine Sushi in PNMA, bi tako lahko bila regulacija programirane celične smrti pri programiranem razvoju tkiv v embrionalnem razvoju. To se sklada z opazovanji, da imajo nekateri geni družine PNMA časovno in prostorsko specifičen vzorec izražanja in bi lahko prek regulacije celične smrti nevronov vplivali na programiran razvoj nevronskih tkiv (Takaji in sod., 2009; Chen in D'Mello, 2010).

### 5.8 BIOLOŠKA VLOGA PROTEINA PNMA5

Biološka vloga večine genov iz družine udomačenih genov PNMA še ni bila določena (pregl. 2). V regulacijo apoptoze so posredno ali neposredno udeleženi geni *PNMA1*, *MOAP1* in *CCDC8* (Tan in sod., 2001; Baksh in sod., 2005; Chen in D'Mello, 2010; Dai in sod., 2011), za protein ZCCHC12 pa je bilo dokazano, da deluje kot transkripcijski ko-aktivator BMP-signalne poti in s tem modulira izražanje genov, specifičnih za holinergične nevrone (Cho in sod., 2009). Gen *PNMA5* ima specifičen vzorec izražanja v asociacijskih področjih neokorteksa pri primatih, njegova

molekularna vloga pa še ni poznana (Takaji in sod., 2009). Dober pokazatelj vpletenosti proteina v posamezne biološke procese je informacija o njegovih vezavnih partnerjih, saj lahko na osnovi poznane biološke vloge molekul v interakciji s preiskovanim proteinom napovemo biološko vlogo tega proteina.

Za identifikacijo vezavnih partnerjev človeškega proteina PNMA5 smo uporabili pristop dvohibridnega sistema kvasovk (Y2H). Na osnovi podatkov o izražanju gena PNMA5 smo pri eksperimentu Y2H uporabili cDNA-knjižnico pripravljeno iz možganskega tkiva človeka. Identificirali smo 5 unikatnih zapisov cDNA, ki kodirajo potencialne vezavne partnerje proteina PNMA5 (pregl. 12). Dva identificirana zapisa kodirata zaporedje človeškega proteina SPTBN1. Identificirana zapisa *SPTBN1* nista prekrivajoča, saj kodirata različna dela proteina SPTBN1 (osrednji in C-končni del proteina) (sl. 15).

Interakcijo med PNMA5 in SPTBN1 smo želeli potrditi s ko-imunoprecipitacijo v lizatu celic SHSY-5Y. Za imunodetekcijo proteinov po prenosu western smo uporabili zajčje protitelo proti SPTBN1, ki se specifično veže na C-konec proteina SPTBN1. Proteina SPTBN1 v frakciji vezanih proteinov nismo zaznali (slika ni prikazana). Interakcije med PNMA5 in SPTBN1 v celicah SHSY-5Y nam ni uspelo potrditi, vendar je za negativen rezultat lahko odgovoren razpad kompleksa PNMA5-SPTBN1 med pripravo celičnih lizatov ali imunoprecipitacijo. Neustrezna je lahko bila tudi izbira uporabljenenih protiteles, saj smo imeli na voljo samo protitelo proti C-koncu proteina SPTBN1, ki se potencialno vežejo s proteinom PNMA5.

Protein SPTBN1 je eden od proteinov iz družine  $\beta$ -spektrinov, ki v spektrinskih molekulah povezujejo aktinski citoskelet in vplivajo na interakcijo aktina s plazemsko membrano (Winkelmann in sod., 1990; Machnicka in sod., 2012). Dokazano je bilo, da se protein SPTBN1 (ELF) pri tretiranju celic s TGF $\beta$ 1 premesti iz citoplazme na citoplazemsko membrano in v jedra celic, se veže s proteini SMAD3 in SMAD4 ter tako vpliva na modulacijo TGF $\beta$ -signaliziranja (Tang in sod., 2003). Modulacija TGF $\beta$ -celičnega signaliziranja prek SPTBN1 (zaradi izgube/zmanjšanega izražanja SPTBN1) se je izkazala za ključen faktor pri nastanku raka gastrointestinalnega trakta in jeter (Katuri in sod., 2005; Baek in sod., 2008; Meindl-Beinker in sod., 2012).

Poleg proteina SPTBN1 je bil v jedru različnih celic identificiran še en protein iz družine ne-eritroidnih  $\beta$ -spektrinov – Sp $\beta$ IV $\Sigma$ 5 (Tse in sod., 2001), ki je 72 kDa velika izooblika proteina SPTBN4 in vključuje spektrinske ponovitve 10-16 proteina polne velikosti (289 kDa). Sp $\beta$ IV $\Sigma$ 5 je v jedru celic ko-lokaliziran s PML jedrnimi telesci (ang. »PML nuclear bodies«), njegova biološka vloga pa ni bila prepoznana (Tse in sod., 2001).

Uporaba fluorescenčnega mikroskopa in konfokalne mikroskopije nam ni podala nedvoumnega odgovora o ko-lokalizaciji proteinov PNMA5 in SPTBN1. Protein SPTBN1 je v celicah SHSY-5Y razpršen tako v citoplazmi kot jedru (sl. 19/A), PNMA5, označen z EGFP (PNMA5-EGFP), pa smo zaznali v obliki točkastih (zrnatih) jedrnih struktur tako pri celicah SHSY-5Y kot pri celicah HEK293T (sl. 17–18).

Vzorec celične lokalizacije proteina PNMA5-EGFP sovpada z vzorcem lokalizacije nekaterih jedrnih struktur, med katere spadajo PML-jedrna telesca. Z imunodetekcijo smo na celičnih preparatih celic HEK293T in SHSY-5Y dokazali, da je protein PNMA5-EGFP delno ko-lokaliziran s PML-jedrnimi telesci (poglavje 4.12). PMLjedrna telesca so majhne jedrne strukture, katerih število na posamezno celico se lahko zelo spreminja (približno 3-30 struktur na celico). Sestava PML-jedrnih telesc je zelo raznolika, saj vsebujejo nekatere stalno prisotne proteine (npr. PML, Sp100 in SUMO) ter številne druge proteine, katerih prisotnost znotraj PML-telesc je odvisna od tipa in stanja celic (Everett in Chelbi-Alix, 2007). Glede na veliko število različnih proteinov znotraj PML-struktur, so bila PML-telesca povezana s številnimi in raznolikimi celičnimi procesi (npr. onkogeneza, stresni odziv, popravljanje DNA, apoptoza, ubikvitinska pot, odziv na IFN, virusni odziv itd.), vendar je o njihovi natančni biološki vlogi težko sklepati (Borden, 2002; Everett in Chelbi-Alix, 2007; Lallemand-Breitenbach in de Thé, 2010). Domneva se, da PML-telesca predstavljajo morfološko oziroma strukturno ogrodje za interakcijo, delovanje in shranjevanje številnih proteinov, udeleženih v različne celične procese znotraj jedra (Borden, 2002).

V literaturi sta opisana še dva primera ko-lokalizacije proteinov, nastalih iz domene gag s PML-jedrnimi telesci. Protein ZCCHC12 iz genske družine PNMA je transkripcijski ko-aktivator BMP-signalne poti. Tako endogeni protein kot z GFP označeni protein ZCCHC12 se ko-lokalizira s PML-jedrnimi strukturami različnih celic (HEK293T, C2C12), pri čemer je prisotnost N-končnega dela proteina, ki vključuje domeno podobno gag, potreben, vendar nezadosten pogoj za pravilno ko-lokalizacijo s PML-telesci (Cho in sod., 2009). Protein ARC, ki je nastal iz domene gag retrotranspozonov, je bil ravno tako identificiran v jedru nevronskih celic in celic HEK293T, kjer je ko-lokaliziran s PML-jedrnimi telesci (Bloomer in sod., 2007). Dodatno potrditev interakcije med ARC in PML-jedrnimi telesci predstavlja identifikacija neposredne interakcije med proteinom ARC in  $\beta$ -spektrinom Sp $\beta$ IV $\Sigma$ 5, za katerega je že bilo pokazano, da interagira s PML-jedrnimi telesci (Tse in sod., 2001).

Interakcija proteina PNMA5 z  $\beta$ -spektrinom SPTBN1, identificirana s pristopom Y2H, nakazuje na morebitno vlogo proteina PNMA5 pri modulaciji TGF $\beta$ 1-signaliziranja prek efektorskih proteinov, kot je npr. SPTBN1. Interakcije med PNMA5 in SPTBN1 nam sicer ni uspelo potrditi, vendar se razlog za neuspeh lahko skriva v uporabi neustreznega protitelesa proti SPTBN1. Z eksperimentom Y2H smo identificirali dva zapisa cDNA, za 2 različna fragmenta proteina SPTBN1 (spektrinske ponovitve 5–8 in

spektrinske ponovitve 13–17 + domena PH), kar nakazuje na morebitno sposobnost vezave proteina PNMA5 in različnih domen spektrinskih ponovitev, ki so lahko vključeni v krajše, še neidentificirane transkripte (izooblike) proteina SPTBN1, kot je to v primeru izooblike proteina SPTBN4 (Sp $\beta$ IV $\Sigma$ 5). Možnost vezave gag-podobnih proteinov s spektrinskimi ponovitvami nakazuje tudi identificirana interakcija proteina ARC s Sp $\beta$ IV $\Sigma$ 5 (Bloomer in sod., 2007).

TGF $\beta$ 1 je citokin s sposobnostjo inhibicije nastanka tumorjev, zmanjša rast celic in prek različnih efektorskih molekul inducira apoptozo. Tumorske celice so, zaradi izgube TGF $\beta$ -signalnih komponent, oziroma zaradi drugih sprememb v povezanih signalnih poteh ali procesih pogosto neodzivne na inhibitorno delovanje TGF $\beta$ 1 (Ikushima in Miyazono, 2010; Massagué, 2012). Invazivnost številnih tumorjev je zaradi sprememb v odzivu na TGF $\beta$ -signaliziranje lahko povečana (Massagué, 2012). Povečano izražanje proteina PNMA5 v celicah HEK293T ima pozitiven vpliv na aktivnost TGF $\beta$ -signalne poti in deluje sinergistično z delovanjem citokina TGF $\beta$ 1 (poglavje 4.14). Iz pridobljenih podatkov lahko sklepamo, da PNMA5 prek interakcije s SPTBN1 pozitivno vpliva na TGF $\beta$ -signalno pot in s tem, odvisno od celičnega okolja – negativno ali pozitivno vpliva na celično proliferacijo.

# 6 SKLEPI

Naredili smo filogenomsko analizo genov, ki so nastali iz ostankov zaporedij Ty3/Gypsy (Metaviridae) naddružine retrotranspozonov. Pridobili smo podatke o številu, razširjenosti, strukturi, kromosomski lokaciji in organiziranosti udomačenih genov na kromosomih.

Prisotnost udomačenih genov smo analizirali pri različnih skupinah placentalnih sesalcev (Boreoeutheria, Afrotheria, Xenarthra), v genomih ključnih vretenčarjev (dvoživke, plazilci, ptiči), ter drugih strunarjev. Široko taksonomsko vzorčenje je bilo ključno za določitev izvora in časa nastanka udomačenih genov in predstavlja bistveno prednost pred predhodnimi študijami.

Dokazali smo, da je večina udomačenih genov prisotna v vseh treh skupinah placentalnih sesalcev (Boreoeutheria, Afrotheria in Xenarthra). Število udomačenih genov je pri drugih sesalcih in pri drugih vretenčarjih bistveno nižje, kar nakazuje na obsežno širitev in diverzifikacijo multigenskih družin udomačenih genov le v predniku placentalnih sesalcev.

S filogenetsko analizo udomačenih genov smo dokazali, da so družine udomačenih genov nastale kot posledica več neodvisnih dogodkov molekularne udomačitve in iz različnih skupin retroelementov družine Ty3/Gypsy.

S filogenetsko analizo retrotranspozonov družine Ty3/Gypsy pri Deuterostomia smo bistveno izboljšali poznavanje razširjenosti teh elementov, še posebno pa pri kopenskih vretenčarjih. S primerjavo razširjenosti udomačenih genov in razširjenosti njihovih izvornih retroelementov smo dokazali, da so udomačeni geni nastali iz ostankov zaporedij retroelementov, ne pa iz aktivnih elementov Ty3/Gypsy.

Filogenomska analiza nam je omogočila pojasnitev para- oziroma mono-filetskega izvora multigenskih družin udomačenih genov. Na podlagi filogenetske analize in strukturne analize genov družine udomačenih genov Sushi sklepamo, da je družina Sushi nastala kot posledica vsaj dveh oziroma treh neodvisnih dogodkov molekularne udomačitve in je torej parafiletskega izvora. V primeru družine PNMA rezultati filogenomske analize kažejo, da je genska družina PNMA monofiletskega izvora.

Glede na strukturno organizacijo elementov Gmr1, ki vključujejo tako SCAN- kot integrazno domeno in na filogenetsko analizo integraznih udomačenih genov sklepamo, da sta gena *KRBA2* in *SCAND3* nastala iz ostankov elementov Gmr1, ne pa s horizontalnim prenosom elementov GINGER2 iz žuželk in naknadno molekularno udomačitvijo integrazne domene teh elementov v predniku Theria, kot je bilo predhodno predlagano (Llorens in sod., 2012).

Analiza kromosomske mobilnosti je dokazala visoko dinamiko nastajanja in premikov udomačenih genov v predniku placentalnih sesalcev, saj smo v primeru genske družine PNMA zaznali številne primere premikov udomačenih genov z avtosomov na kromosom X, ter z kromosoma X na avtosomalne kromosome.

Dokazali smo, da so nukleotidna in aminokislinska zaporedja udomačenih genov in njihovi kromosomski položaji pri placentalnih sesalcih izjemno ohranjeni. Zelo dinamično nastajanje novih udomačenih genov in hitra fiksacija teh genov pred radiacijo placentalnih sesalcev nakazuje na močno adaptivno evolucijo in neofunkcionalizacijo teh zaporedij v predniku placentalnih sesalcev.

Pojasnili smo ključne molekularne dogodke, ki so vodili od zaporedij nekoč aktivnih retroelementov do nastanka novih funkcionalnih genov in s tem bistveno izboljšali poznavanje procesa molekularne domestikacije pri sesalcih.

Dokazali smo, da so udomačeni geni nastali iz ostankov retroelementov, ki nimajo lastnih promotorskih zaporedij, oziroma se njihove regulatorne regije (LTR) niso ohranile, zato lahko sklepamo, da so udomačeni geni svoja promotorska in druga regulatorna zaporedja pridobili na novo.

Analiza promotorskih zaporedij udomačenih genov je dokazala, da so udomačeni geni pogosto pridobili promotorska zaporedja bogata s CpG-dinukleotidi, ki prej niso bila del promotorjev obstoječih genov. Dostop do oddaljenih promotorjev je udomačenim genom omogočila evolucija intronsko/eksonskih struktur njihovih na novo pridobljenih 5'-neprevedenih regij.

Analiza dolžin 3'-neprevedenih zaporedij udomačenih genov je dokazala, da so te regije pri večini udomačenih genov nenavadno dolge ter bi lahko vplivale na stabilnost mRNA in regulacijo izražanja udomačenih genov.

Filogenomska analiza udomačenih genov je skupaj z analizo literaturnih podatkov pokazala, da so imeli udomačeni geni ključno vlogo pri pojavu in razvoju pomembnih fenotipskih novosti pri placentalnih sesalcih, kot sta npr. razvoj placente in povečanje velikosti/kompleksnosti možganov.

S pomočjo Y2H smo identificirali vezavne partnerje proteina PNMA5. Interakcija med SPTBN1 in PNMA5, določena z Y2H, nakazuje morebitno vlogo proteina PNMA5 pri modulaciji TGFβ-celičnega signaliziranja.

Dokazali smo, da izražanje proteina PNMA5 v celicah HEK293T deluje sinergistično z delovanjem citokina TGFβ1 in pozitivno modulira aktivnost TGFβ-signalne poti. Dokazali smo tudi, da se protein PNMA5 nahaja v jedru celic HEK293T in SHSY-5Y, kjer interagira s PML-jedrnimi telesci.

Celovit pogled na nastanek in evolucijo udomačenih genov, nastalih iz retroelementov, kaže, da je imel pojav udomačenih genov, vključno s pridobivanjem novih regulatornih regij (regulatorno ožičenje genov), pomemben vpliv na razvoj in pridobivanje evolucijskih (fenotipskih) novosti pri placentalnih sesalcih.

# 7 POVZETEK (SUMMARY)

## 7.1 POVZETEK

Transpozicijski elementi (TE) so mobilni genetski elementi, ki imajo sposobnost premikanja in replikacije znotraj genoma gostitelja (Kidwell in Lisch, 2000; Kidwell, 2001; Kazazian, 2004; Goodier, 2008). TE s svojim delovanjem vplivajo na številne celične procese ter so pomembno sodelovali pri evoluciji in preoblikovanju vretenčarskih genomov, hkrati pa so vedno bolj prepoznani kot pomemben vir novih kodirajočih zaporedij v obliki eksonov in celotnih genov, s katerimi so obogatili gostiteljev genski nabor (Volff, 2006; Böhne in sod., 2008, Kaessmann, 2010). Genomi vretenčarjev, predvsem pa sesalcev, vsebujejo številna kodirajoča zaporedja, ki so nastala iz TE (Zdobnov in sod., 2005; Campillos in sod., 2006, Volff, 2006). Zaporedia TE, ki kodirajo različne zapise gag, pol in env, so v procesu molekularne udomačitve pogosto pridobila sposobnost kodiranja novih funkcionalnih proteinov in pridobila nove biološke vloge (postala so neofunkcionalizirana) (Miller in sod., 1999; Zdobnov in sod., 2005, Volff, 2006). Število prepoznanih bioloških vlog pri udomačenih genih nakazuje, da so domene retrotranspozonov visoko prilagodljivi interakcijski moduli, prek katerih lahko interagirajo različne biološke molekule (nukleinske kisline in proteini) (Campillos et al. 2006; Volff 2006; Feschotte 2008). Novonastali geni lahko nove biološke funkcije pridobijo v procesu adaptivne evolucije, oziroma prek evolucije njihovih regulatornih regij, s čimer razvijejo tkivno in/ali časovno odvisen vzorec izražanja (Volff, 2006).

Nastanek in evolucija udomačenih genov, ki so nastali iz družine retroelementov Ty3/Gypsy (Metaviridae), je bila do sedaj le delno pojasnjena. Pomanjkljivost obstoječih študij je bilo predvsem omejeno taksonomsko vzorčenje (Lloréns in Marín, 2001; Zdobnov in sod., 2005; Campillos in sod., 2006), oziroma je bilo zadovoljivo vzorčenje uporabljeno samo za eno družino udomačenih genov (Lynch in sod., 2003; Brandt in sod., 2005b). Obstoječe študije tako niso podale celovitega vpogleda v nastanek, razširjenost in evolucijo genov, ki so nastali iz retroelementov.

Naredili smo filogenomsko analizo genov, ki so nastali iz zaporedij družine retrotranspozonov Ty3/Gypsy. Pridobili smo podatke o številu, razširjenosti, strukturi, kromosomski lokaciji in organiziranosti udomačenih genov na kromosomih. Pojasnili smo evolucijske odnose znotraj multigenskih družin udomačenih genov in odgovorili na vprašanja o para-oziroma mono-filetskem izvoru teh družin (Brandt in sod., 2005a). Prisotnost udomačenih genov smo analizirali pri različnih skupinah placentalnih sesalcev (Boreoeutheria, Afrotheria, Xenarthra), v genomih ključnih vretenčarjev (dvoživke, plazilci, ptiči) in drugih strunarjev. Široko taksonomsko vzorčenje je bilo ključno za določitev izvora in časa nastanka udomačenih genov.

Dokazali smo, da so različne družine udomačenih genov nastale neodvisno, iz ostankov različnih skupin retroelementov družine Ty3/Gypsy. Široka razširjenost teh genov v vseh štirih nadredih placentalnih sesalcev nakazuje na obsežno širitev in diverzifikacijo multigenskih družin udomačenih genov v predniku placentalnih sesalcev (Eutheria). Zaporedja in kromosomski položaji udomačenih genov so po ločitvi štirih placentalnih nadredov ostali močno ohranjeni, kar kaže na močno adaptivno evolucijo in neofunkcionalizacijo teh genov.

Dokazali smo, da so udomačeni geni nastali iz zaporedij ostankov retroelementov, ki nimajo lastnih promotorskih zaporedij oziroma se njihove regulatorne regije (LTR) niso ohranile, zato lahko sklepamo, da so udomačeni geni svoja promotorska, 3'- in 5'- neprevedena zaporedja pridobili na novo. Pridobitev in evolucija regulatornih zaporedij udomačenih genov sta igrali pomembno vlogo pri adaptivni evoluciji teh genov ter sta pomembni za razlago opaženih tkivno, vrstno in časovno odvisnih vzorcev izražanja teh genov.

Gen *PNMA5* iz multigenske družine PNMA ima značilen vzorec izražanja v možganih primatov (Takaji in sod., 2009), zato nas je zanimala njegova biološka vloga. Z Y2H smo mu določili vezavne partnerje. Dokazali smo, da se PNMA5 v jedru celic HEK293T in SHSY-5Y povezuje s PML-jedrnimi strukturami. Rezultati Y2H in luciferaznega testa kažejo, da protein PNMA5 – verjetno prek interakcije s proteinom SPTBN1 – pozitivno modulira aktivnost TGFβ1 signalne poti.

Filogenomska analiza udomačenih genov, skupaj z analizo literaturnih podatkov kaže, da so ti geni imeli ključno vlogo pri nastanku in razvoju pomembnih fenotipskih novosti pri placentalnih sesalcih, kot sta npr. razvoj placente in povečanje velikosti/kompleksnosti možganov.

### 7.2 SUMMARY

Transposable elements (TEs) constitute a major component of eukaryotic genomes and have profound effects on the structure, function and evolution of their host genomes (Kidwell and Lisch, 2000; Kidwell, 2001; Kazazian, 2004; Goodier, 2008). The mobility and amplification of TEs represents a major source of genomic variation, by virtue of their insertion or by triggering a variety of small- and large-scale chromosomal rearrangements. During evolution, many novel protein-coding genes have emerged from the long terminal repeat (LTR) retroelements (retrotransposons and retroviruses). Different retrotransposon domains, including gag, envelope, integrase and protease genes, frequently became exapted and encode novel (domesticated) genes or exons (Campillos et al. 2006; Volff, 2006, Böhne in sod., 2008, Kaessmann, 2010). Exaptation or molecular domestication of TE-derived sequences is a complex molecular process involving many changes at the nucleotide level and can lead, through adaptive evolution

of coding and regulatory regions, to novel genes that can be beneficial for their host (Miller et al., 1999; Zdobnov et al., 2005, Volff, 2006).

The evolutionary history and dynamics of novel Ty3/Gypsy (Metaviridae) derived domesticated genes have been only partially explored, due to absence of genome data or to the limited analysis of a single family of domesticated genes (Brandt et al. 2005a; Brandt et al. 2005b; Lynch and Tristem 2003; Gorinšek et al. 2004; Zdobnov et al. 2005; Campillos et al. 2006; Llorens and Marin 2001). The origin and evolution of different retrotransposon-derived gene families was therefore not adequately explained. By using much wider taxon sampling, we provided new insight into the origin of domesticated genes, the timing of molecular domestication events, and the evolution of these multi-gene families.

By using phylogenomic analysis we obtained and characterized domesticated genes and their Metaviridae progenitors from all currently available mammalian (>50 different species available at NCBI) genomes, key tetrapod (amphibians and reptiles) and the remaining chordate genomes. In total, more than 90 chordate genomes were analyzed. Phylogenomic analysis provided a large amount of information for any novel retrotransposon-derived domesticated gene, including genome sequence, gene structure, chromosome location, protein sequence, coding and non-coding regions, as well as regulatory regions.

Phylogenetic and sequence analysis of retroelement-derived domesticated genes provides strong evidence that they originated independently several times in the ancestor of placentals.

We traced the progenitors of domesticated genes in amniotes. This approach enabled us to show unequivocally that the gag- and integrase-derived domesticated genes originated from Metaviridae remains (molecular fossils), since no active Chromovirus or Barthez lineages of Metaviridae are present in any mammalian genome.

By using phylogenomic analysis, we explained where and when molecular domestication (exaptation) of retroelement-derived sequences occurred. We showed that majority of domesticated genes originated in the ancestor of placental mammals and that domesticated genes and their chromosomal positions were fully established in the ancestor of eutherian mammals. We demonstrated that strong adaptive evolution had diversified domesticated genes before their fixation in the ancestor of placental mammals. After fixation, domesticated genes evolved under strong purifying selection (Brandt et al., 2005a).

Since gag, protease and integrase domains in retroelements lack promoters and UTRs, they must have been acquired *de novo* in domesticated genes. We have analyzed the origin and evolution of *de novo* acquired promoters, 5' UTRs and 3' UTRs in diverse

mammalian domesticated genes by comparative analysis of orthologous gene loci. The acquisition of promoter and regulatory regions by domesticated genes was crucial for the survival of these retrotransposon-derived sequences.

It has been previously suggested that PNMA5 gained specialized role in the association areas of the neocortex during primate evolution (Takaji et al., 2009). To elucidate its biological role, we performed "Yeast Two Hybrid" (Y2H) screen to identify its binding partners. The interaction between PNMA5 protein and the major actin crosslinking protein SPTBN1 has been identified. GFP-tagged human PNMA5 protein expressed in the SHSY-5Y cells was localized in the nucleus in speckled pattern. Immunofluorescence staining further revealed its co-localization with the PML nuclear bodies. Protein interaction data and sub-cellular localization analysis provided the clue about possible role of the PNMA5 protein in the modulation of the TGF $\beta$ 1 cell signaling in the neuronal cells.

Retroelement-derived domesticated genes have, through adaptive evolution of their coding and regulatory regions, frequently acquired novel biological roles (they became neofunctionalized). The origin of placental mammals specific innovations and adaptations, such as placenta and newly evolved brain functions, was most probably connected to the regulatory wiring of domesticated genes and their rapid fixation in the ancestor of placental mammals.
# 8 VIRI

- Alföldi J., Di Palma F., Grabherr M., Williams C., Kong L., Mauceli E., Russell P., Lowe C.B., Glor R.E., Jaffe J.D., Ray D.A., Boissinot S., Shedlock A.M., Botka C., Castoe T.A., Colbourne J.K., Fujita M.K., Moreno R.G., ten Hallers B.F., Haussler D., Heger A., Heiman D., Janes D.E., Johnson J., de Jong P.J., Koriabine M.Y., Lara M., Novick P.A., Organ C.L., Peach S.E., Poe S., Pollock D.D., de Queiroz K., Sanger T., Searle S., Smith J.D., Smith Z., Swofford R., Turner-Maier J., Wade J., Young S., Zadissa A., Edwards S.V., Glenn T.C., Schneider C.J., Losos J.B., Lander E.S., Breen M., Ponting C.P., Lindblad-Toh K. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature, 477, 7366: 587–591
- Anderson J. S. 2008. Focal review: The origin (s) of modern amphibians. Evolutionary Biology, 35, 4: 231-247
- Andreassi C., Riccio A. 2009. To localize or not to localize: mRNA fate is in 3'UTR ends. Trends in Cell Biology, 19, 9: 465-474
- Baek H. J., Lim S. C., Kitisin K., Jogunoori W., Tang Y., Marshall M. B., Mishra B., Kim T.H., Cho K.H., Kim S.S. 2008. Hepatocellular cancer arises from loss of transforming growth factor beta signaling adaptor protein embryonic liver fodrin through abnormal angiogenesis. Hepatology, 48, 4: 1128-1137
- Baksh S., Tommasi S., Fenton S., Yu V.C., Martins L.M., Pfeifer G.P., Latif F., Downward J., Neel B.G. 2005. The tumor suppressor RASSF1A and MAP-1 link death receptor signaling to Bax conformational change and cell death. Molecular Cell, 18, 6: 637-650
- Barlow D. P. 1993. Methylation and imprinting: from host defense to gene regulation? Science, 260, 5106: 309-310
- Barrett L. W., Fletcher S., Wilton S. D. 2012. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cellular and Molecular Life Sciences, 69, 21: 3613-3634
- Biémont C., Vieira C. 2006. Genetics: junk DNA as an evolutionary force. Nature, 443, 7111: 521-524
- Blaise S., de Parseval N., Benit L., Heidmann T. 2003. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. Proceedings of the National Academy of Sciences of the United States of America, 100, 22: 13013–13018

- Blaise S., de Parseval N., Heidmann T. 2005. Functional characterization of two newly identified Human Endogenous Retrovirus coding envelope genes. Retrovirology 2, 1: 19 doi:10.1186/1742-4690-2-19
- Bloomer W. A., VanDongen H., VanDongen A. M. 2007. Activity-regulated cytoskeleton-associated protein Arc/Arg3.1 binds to spectrin and associates with nuclear promyelocytic leukemia (PML) bodies. Brain Research, 1153: 20-33
- Bock C., Walter J., Paulsen M., Lengauer T. 2007. CpG island mapping by epigenome prediction. PLoS Computational Biology, 3, 6 : e110 doi:10.1371/journal.pcbi.0030110
- Borden K. L. 2002. Pondering the promyelocytic leukemia protein (PML) puzzle: possible functions for PML nuclear bodies. Molecular and Cellular Biology, 22, 15: 5259-5269
- Bourque G., Leong B., Vega V.B., Chen X., Lee Y.L., Srinivasan K.G., Chew J.L., Ruan Y., Wei C.L., Ng H.H., Liu E.T. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Research, 18, 11: 1752-1762
- Brandt J., Schrauth S., Veith A.M., Froschauer A., Haneke T., Schultheis C., Gessler M., Leimeister C., Volff J.N. 2005a. Transposable elements as a source of genetic innovation: expression and evolution of a family of retrotransposon-derived neogenes in mammals. Gene, 345, 1: 101-111
- Brandt J., Veith A.M., Volff J.N. 2005b. A family of neofunctionalized Ty3/gypsy retrotransposon genes in mammalian genomes. Cytogenetic and Genome Research, 110: 307-317
- Brosius J. 1991. Retroposons-seeds of evolution. Science, 251, 4995: 753
- Brosius J. 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene, 238, 1: 115-134
- Butler M., Goodwin T., Poulter R. 2001. An unusual vertebrate LTR retrotransposon from the cod Gadus morhua. Molecular Biology and Evolution, 18, 3: 443-447
- Böhne A., Brunet F., Galiana-Arnoux D., Schultheis C., Volff JN. 2008. Transposable elements as drivers of genomic and biological diversity in vertebrates. Chromosome Research, 16, 1: 203-215

- Campillos M., Doerks T., Shah P.K., Bork P. 2006. Computational characterization of multiple Gag-like human proteins. Trends in Genetics, 22, 11: 585-589
- Capy P., Gasperi G., Biémont C., Bazin C. 2000. Stress and transposable elements: coevolution or useful parasites? Heredity, 85, 2: 101-106
- Chalopin D., Galiana D., Volff J.N. 2012. Genetic innovation in vertebrates: gypsy integrase genes and other genes derived from transposable elements. International Journal of Evolutionary Biology, 2012: 724519 doi:10.1155/2012/724519
- Charlier C., Segers K., Wagenaar D., Karim L., Berghmans S., Jaillon O., Shay T., Weissenbach J., Cockett N., Gyapay G., Georges M. 2001. Human-ovine comparative sequencing of a 250-kb imprinted domain encompassing the callipyge (clpg) locus and identification of six imprinted transcripts: DLK1, DAT, GTL2, PEG11, antiPEG11, and MEG8. Genome Research, 11, 5: 850–862
- Chen F.C., Li W.H. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. American Journal of Human Genetics, 68, 2: 444-456
- Chen H.L., D'Mello S.R., 2010. Induction of neuronal cell death by paraneoplastic Ma1 antigen. Journal of Neuroscience Research, 88, 16: 3508-3519
- Cheung H.H., Lee T.L., Davis A.J., Taft D.H., Rennert O.M., Chan W.Y. 2010. Genome-wide DNA methylation profiling reveals novel epigenetically regulated genes and non-coding RNAs in human testicular cancer. British Journal of Cancer, 102, 2: 419-427
- Cho G., Lim Y., Zand D., Golden J.A. 2008. Sizn1 is a novel protein that functions as a transcriptional coactivator of bone morphogenic protein signaling. Molecular Cell Biology, 28, 5: 1565-1572
- Cho G., Lim Y., Golden J.A. 2009. SUMO interaction motifs in Sizn1 are required for promyelocytic leukemia protein nuclear body localization and for transcriptional activation. Journal of Biological Chemistry, 284, 29: 19592-19600
- Chowdhury S., Shepherd J.D., Okuno H., Lyford G., Petralia R.S., Plath N., Kuhl D., Huganir R.L., Worley P.F. 2006. Arc/Arg3.1 interacts with the endocytic machinery to regulate AMPA receptor trafficking. Neuron, 52, 3: 445-459
- Churakov G., Kriegs J.O., Baertsch R., Zemann A., Brosius J., Schmitz J. 2009. Mosaic retroposon insertion patterns in placental mammals. Genome Research, 19, 5: 868-875

- Clark M.B., Jänicke M., Gottesbühren U., Kleffmann T., Legge M., Poole E.S., Tate W.P. 2007. Mammalian gene PEG10 expresses two reading frames by high efficiency -1 frameshifting in embryonic-associated tissues. Journal of Biological Chemistry, 282, 52: 37359-37369
- Clack J.A. 2012. Gaining ground: the origin and evolution of tetrapods. 2. izd. Bloomington, Indiana, USA, Indiana University Press: 369 str.
- Cornelis G., Heidmann O., Bernard-Stoecklin S., Reynaud K., Véron G., Mulot B., Dupressoir A., Heidmann T 2012. Ancestral capture of syncytin-Car1, a fusogenic endogenous retroviral envelope gene involved in placentation and conserved in Carnivora. Proceedings of the National Academy of Sciences of the United States of America, 109, 7: E432-E441 doi:10.1073/pnas.1115346109
- Dai C., Tang Y., Jung S.Y., Qin J., Aaronson S.A., Gu W. 2011. Differential effects on p53-mediated cell cycle arrest vs. apoptosis by p90. Proceedings of the National Academy of Sciences of the United States of America, 108, 47: 18937-18942
- Davis E., Caiment F., Tordoir X., Cavaillé J., Ferguson-Smith A., Cockett N., Georges M., Charlier C. 2005. RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. Current Biology, 15, 8: 743-749
- Doolittle W.F., Sapienza C., 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature, 284, 5757: 601-603
- Dupressoir A., Marceau G., Vernochet C., Bénit L., Kanellopoulos C., Sapin V., Heidmann T. 2005. Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. Proceedings of the National Academy of Sciences of the United States of America, 102, 3: 725– 730
- Dupressoir A., Vernochet C., Harper F., Guégan J., Dessen P., Pierron G., Heidmann T. 2011. A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast. Proceedings of the National Academy of Sciences of the United States of America, 108, 46: E1164-E1173 doi:10.1073/pnas.1112304108
- Edelstein L.C., Collins T. 2005. The SCAN domain family of zinc finger transcription factors. Gene, 359: 1–17

- Eissenberg J.C. 2001. Molecular biology of the chromo domain: an ancient chromatin module comes of age. Gene, 275, 1: 19-29
- Emerson R.O., Thomas J.H., 2011. Gypsy and the birth of the SCAN domain. Journal of Virology, 85, 22: 12043–12052
- Esnault C., Maestre J., Heidmann T. 2000. Human LINE retrotransposons generate processed pseudogenes. Nature Genetics, 24, 4: 363–367
- Fablet M., Bueno M., Potrzebowski L., Kaessmann H. 2009. Evolutionary origin and functions of retrogene introns. Molecular Biology and Evolution, 26, 9: 2147– 2156
- Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. Nature Reviews Genetics, 9, 5: 397-405
- Fields S., Song O. 1989. A novel genetic system to detect protein-protein interactions. Nature, 340: 245-246
- Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32, 5: 1792-1797
- Everett R.D., Chelbi-Alix, M.K. 2007. PML and PML nuclear bodies: implications in antiviral defence. Biochimie, 89, 6-7: 819
- Fu N.Y., Sukumaran S.K., Yu V.C. 2007. Inhibition of ubiquitin-mediated degradation of MOAP-1 by apoptotic stimuli promotes Bax function in mitochondria. Proceedings of the National Academy of Sciences of the United States of America, 104, 24: 10050-10056
- Gertz E.M., Yu Y.K., Agarwala R., Schäffer A.A., Altschul S.F. 2006. Compositionbased statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biology, 4, 1: 41 doi:10.1186/1741-7007-4-41
- Goodier J.L., Kazazian H.H. Jr. 2008. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell, 135, 1: 23-35
- Goodwin T.J., Poulter R.T., 2002. A group of deuterostome Ty3/gypsy-like retrotransposons with Ty1/copia-like pol-domain orders. Molecular Genetics and Genomics, 267, 4: 481–491
- Gould S.J., Vrba E.S. 1982. Exaptation: a missing term in the science of form. Paleobiology, 8, 1: 4–15

- Gorinšek B., Gubenšek F., Kordiš D., 2004. Evolutionary genomics of chromoviruses in eukaryotes. Molecular Biology and Evolution, 21, 5: 781-798
- Gorinšek B., Gubenšek F., Kordiš, D. 2005. Phylogenomic analysis of chromoviruses. Cytogenetic and Genome Research, 110, 1-4: 543-552
- Grillo G., Turi A., Licciulli F., Mignone F., Liuni S., Banfi S., Gennarino V.A., Horner D.S., Pavesi G., Picardi E., Pesole G. 2010. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Research, 38, suppl 1: D75-D80 doi:10.1093/nar/gkp902
- Guindon S., Lethiec F., Duroux P., Gascuel O. 2005. PHYML Online a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Research, 33, suppl 2: W557-W559 doi:10.1093/nar/gki352
- Hanson D., Murray P.G., Black G.C., Clayton P.E. 2011. The genetics of 3-M syndrome: unravelling a potential new regulatory growth pathway. Hormone Research in Pædiatrics, 76, 6: 369-378
- Hedges S.B. 2009. Vertebrates (Vertebrata). V: The Timetree of Life. Hedges S. B., Kumar S. (eds.). New York, Oxford University Press: 309-314
- Hill R.S., Walsh C.A. 2005. Molecular insights into human brain evolution. Nature, 437, 7055: 64-67
- Huelsenbeck J.P., Ronquist F. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics, 17, 8: 754-755
- Ikushima H., Miyazono K. 2010. TGFβ signalling: a complex web in cancer progression. Nature Reviews Cancer, 10, 6: 415-424
- Innan H., Kondrashov F., 2010. The evolution of gene duplications: Classifying and distinguishing between models. Nature Reviews Genetics, 11, 2: 97–108
- Inoue M., Takahashi K., Niide O., Shibata M., Fukuzawa M., Ra C. 2005. LDOC1, a novel MZF-1-interacting protein, induces apoptosis. FEBS Letters, 579, 3: 604-608
- Jurka J. 2004. Evolutionary impact of human Alu repetitive elements. Current Opinion in Genetics and Development, 14, 6: 603-608

- Jurka J., Kapitonov V.V., Kohany O., Jurka M.V. 2007. Repetitive Sequences in Complex Genomes. Structure and Evolution. Annual Review of Genomics and Human Genetics, 8: 241-259
- Kaessmann H., Vinckenbosch N., Long M. 2009. RNA-based gene duplication: Mechanistic and evolutionary insights. Nature Reviews Genetics, 10, 1: 19–31
- Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. Genome Research, 20, 10: 1313-1326
- Kalitsis P., Saffery R. 2009. Inherent promoter bidirectionality facilitates maintenance of sequence integrity and transcription of parasitic DNA in mammalian genomes. BMC Genomics, 10, 1: 498 doi:10.1186/1471-2164-10-498
- Kaneko-Ishino T., Kohda T., Ono R., Ishino F. 2006. Complementation hypothesis: the necessity of a monoallelic gene expression mechanism in mammalian development. Cytogenetics Genome Research, 113, 1-4: 24-30
- Kaneko-Ishino T., Ishino F., 2012. Evolution of Viviparity and Genomic Imprinting in Mammals by Retrotransposons. V: Evolutionary Biology: Mechanisms and Trends. Pontarotti P. (ed.). Berlin Heidelberg, Springer: 265-281 doi:10.1007/978-3-642-30425-5\_15
- Kapitonov V.V., Jurka J., 2005. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biology, 3, 6: e181 doi:10.1371/journal.pbio.0030181
- Kapitonov V.V., Jurka J., 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. Nature Reviews Genetics, 9, 5: 411-412
- Katuri V., Tang Y., Marshall B., Rashid A., Jogunoori W., Volpe E.A., Sidawy A.N, Evans S., Blay J., Gallicano G.I., Premkumar Reddy E., Mishra L. 2005. Inactivation of ELF/TGF-β signaling in human gastrointestinal cancer. Oncogene, 24, 54: 8012-8024
- Kazazian H. H. 2004. Mobile Elements: Drivers of Genome Evolution. Science 303, 5664: 1626-1632
- Khil P. P., Smirnova N. A., Romanienko P. J., Camerini-Otero R. D. 2004. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. Nature Genetics, 36, 6: 642-646

- Kidwell M.G., Lisch D.R. 2000. Transposable elements and host genome evolution. Trends in Ecology and Evolution, 15, 3: 95-99
- Kidwell M.G., 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica, 115, 1: 49-63
- Kidwell M.G., Lisch D.R., 2001. Perspective: Transposable elements, parasitic DNA and genome evolution. Evolution, 55, 1: 1-24
- Korb E., Finkbeiner S. 2011. Arc in synaptic plasticity: from gene to behavior. Trends Neuroscience, 34, 11: 591-598
- Kordiš D., 2005. A genomic perspective on the chromodomain-containing retrotransposons: Chromoviruses. Gene, 347, 2: 161-173
- Kordiš D., Lovsin N., Gubensek F., 2006. Phylogenomic analysis of the L1 retrotransposons in Deuterostomia. Systematic Biology, 55, 6: 886-901
- Kordiš D. 2009. Transposable elements in reptilian and avian (sauropsida) genomes. Cytogenetic and Genome Research, 127, 2-4: 94–111
- Kordiš D. 2011. Extensive intron gain in the ancestor of placental mammals. Biology Direct, 6: 59 doi:10.1186/1745-6150-6-59
- Kordiš D., Kokošar J. 2012. What Can Domesticated Genes Tell Us about the Intron Gain in Mammals? International Journal of Evolutionary Biology, 2012: 278981 doi:10.1155/2012/278981
- Krull M., Brosius J., Schmitz J. 2005. Alu-SINE exonization: En route to proteincoding function. Molecular Biology and Evolution, 22, 8: 1702–1711
- Laketa V., Simpson J.C., Bechtel S., Wiemann S., Pepperkok R. 2007. High-content microscopy identifies new neurite outgrowth regulators. Molecular Biology of the Cell, 18, 1: 242-252
- Lallemand-Breitenbach V., de Thé H. 2010. PML nuclear bodies. Cold Spring Harbor Perspectives in Biology, 2, 5: a000661 doi:10.1101/cshperspect.a000661
- Lander, E. S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., Funke R., Gage D., Harris K., Heaford A., Howland J., Kann L., Lehoczky J., LeVine R., McEwan P., McKernan K., Meldrim J., Mesirov J.P., Miranda C., Morris W., Naylor J., Raymond C., Rosetti

M., Santos R., Sheridan A., Sougnez C., Stange-Thomann N., Stojanovic N., Subramanian A., Wyman D., Rogers J., Sulston J., Ainscough R., Beck S., Bentley D., Burton J., Clee C., Carter N., Coulson A., Deadman R., Deloukas P., Dunham A., Dunham I., Durbin R., French L., Grafham D., Gregory S., Hubbard T., Humphray S., Hunt A., Jones M., Lloyd C., McMurray A., Matthews L., Mercer S., Milne S., Mullikin J.C., Mungall A., Plumb R., Ross M., Shownkeen R., Sims S., Waterston R.H., Wilson R.K., Hillier L.W., McPherson J.D., Marra M.A., Mardis E.R., Fulton L.A., Chinwalla A.T., Pepin K.H., Gish W.R., Chissoe S.L., Wendl M.C., Delehaunty K.D., Miner T.L., Delehaunty A., Kramer J.B., Cook L.L., Fulton R.S., Johnson D.L., Minx P.J., Clifton S.W., Hawkins T., Branscomb E., Predki P., Richardson .P, Wenning S., Slezak T., Doggett N., Cheng J.F., Olsen A., Lucas S., Elkin C., Uberbacher E., Frazier M., Gibbs R.A., Muzny D.M., Scherer S.E., Bouck J.B., Sodergren E.J., Worley K.C., Rives C.M., Gorrell J.H., Metzker M.L., Naylor S.L., Kucherlapati R.S., Nelson D.L., Weinstock G.M., Sakaki Y., Fujiyama A., Hattori M., Yada T., Toyoda A., Itoh T., Kawagoe C., Watanabe H., Totoki Y., Taylor T., Weissenbach J., Heilig R., Saurin W., Artiguenave F., Brottier P., Bruls T., Pelletier E., Robert C., Wincker P., Smith D.R., Doucette-Stamm L., Rubenfield M., Weinstock K., Lee H.M., Dubois J., Rosenthal A., Platzer M., Nyakatura G., Taudien S., Rump A., Yang H., Yu J., Wang J., Huang G., Gu J., Hood L., Rowen L., Madan A., Qin S., Davis R.W., Federspiel N.A., Abola A.P., Proctor M.J., Myers R.M., Schmutz J., Dickson M., Grimwood J., Cox D.R., Olson M.V., Kaul R., Raymond C., Shimizu N., Kawasaki K., Minoshima S., Evans G.A., Athanasiou M., Schultz R., Roe B.A., Chen F., Pan H., Ramser J., Lehrach H., Reinhardt R., McCombie W.R., de la Bastide M., Dedhia N., Blöcker H., Hornischer K., Nordsiek G., Agarwala R., Aravind L, Bailey J.A., Bateman A., Batzoglou S., Birney E., Bork P., Brown D.G., Burge C.B., Cerutti L., Chen H.C., Church D., Clamp M., Copley R.R., Doerks T., Eddy S.R., Eichler E.E., Furey T.S., Galagan J., Gilbert J.G., Harmon C., Hayashizaki Y., Haussler D., Hermjakob H., Hokamp K., Jang W., Johnson L.S., Jones T.A., Kasif S., Kaspryzk A., Kennedy S., Kent W.J., Kitts P., Koonin E.V., Korf I., Kulp D., Lancet D., Lowe T.M., McLysaght A., Mikkelsen T., Moran J.V., Mulder N., Pollara V.J., Ponting C.P., Schuler G., Schultz J., Slater G., Smit A.F., Stupka E., Szustakowski J., Thierry-Mieg D., Thierry-Mieg J., Wagner L., Wallis J., Wheeler R., Williams A., Wolf Y.I., Wolfe K.H., Yang S.P., Yeh R.F., Collins F., Guyer M.S., Peterson J., Felsenfeld A., Wetterstrand K.A., Patrinos A., Morgan M.J., de Jong P., Catanese J.J., Osoegawa K., Shizuya H., Choi S., Chen Y.J.; International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature, 409, 6822: 860-921

Lev-Maor G., Sorek R., Shomron N., Ast G. 2003. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. Science, 300, 5623: 1288–1291

- Lev-Maor G., Sorek R., Levanon E.Y., Paz N., Eisenberg E., Ast G., 2007. RNAediting-mediated exon evolution. Genome Biology, 8, 2: R29 doi:10.1186/gb-2007-8-2-r29
- Lloréns C., Marín I. 2001. A mammalian gene evolved from the integrase domain of an LTR retrotransposon. Molecular Biology and Evolution, 18, 8: 1597–1600
- Llorens C., Bernet G., Ramasamy S., Feschotte C., Moya A. 2012. On the transposon origins of mammalian SCAND3 and KRBA2, two zinc-finger genes carrying an integrase/transposase domain. Mobile Genetic Elements, 2, 5: 205-210
- Long M., Betran E., Thornton K., Wang W. 2003. The origin of new genes: glimpses from the young and old. Nature Reviews Genetics, 4, 11: 865–875
- Luo Z.X., Yuan C.X., Meng Q.J., Ji Q. 2011. A Jurassic eutherian mammal and divergence of marsupials and placentals. Nature, 476, 7361: 442-445
- Lux A., Beil C., Majety M., Barron S., Gallione C.J., Kuhn H.M., Berg J.N., Kioschis P., Marchuk D.A., Hafner M. 2005. Human retroviral gag- and gag-pol-like proteins interact with the transforming growth factor-beta receptor activin receptor-like kinase 1. Journal of Biological Chemistry, 280, 9: 8482-8493
- Lyford G.L., Yamagata K., Kaufmann W.E., Barnes C.A., Sanders L.K., Copeland N.G., Gilbert D.J., Jenkins N.A., Lanahan A.A., Worley P.F. 2005. Arc, a growth factor and activity-regulated gene, encodes a novel cytoskeleton-associated protein that is enriched in neuronal dendrites. Neuron, 14, 2: 433-445
- Lynch C., Tristem M. 2003. A co-opted gypsy-type LTR-retrotransposon is conserved in the genomes of humans, sheep, mice, and rats. Current Biology, 13, 17: 1518-1523
- Machnicka B., Grochowalska R., Bogusławska D. M., Sikorski A. F., Lecomte, M. C. 2012. Spectrin-based skeleton as an actor in cell signaling. Cellular and Molecular Life Sciences, 69, 2: 191-201
- Madsen O. 2009. Mammals (Mammalia). V: The Timetree of Life. Hedges S. B., Kumar S. (eds.). New York, Oxford University Press: 459-661
- Malik H.S., Eickbush T.H. 1999. Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. Journal of Virology, 73, 6: 5186-5190
- Marco A., Marín I. 2009. CGIN1: a retroviral contribution to mammalian genomes. Molecular Biology and Evolution, 26, 10: 2167–2170

- Marín I., 2010. GIN transposons: genetic elements linking retrotransposons and genes. Molecular Biology and Evolution, 27, 8: 1903–1911
- Massagué J. 2012. TGFβ signalling in context. Nature Reviews Molecular Cell Biology, 13: 616-630
- Matsui T., Kinoshita-Ida Y., Hayashi-Kisumi F., Hata M., Matsubara K., Chiba M., Katahira-Tayama S., Morita K., Miyachi Y., Tsukita S. 2006. Mouse homologue of skin-specific retroviral-like aspartic protease involved in wrinkle formation. Journal of Biological Chemistry, 281, 37: 27512–27525
- Matsui T., Miyamoto K., Kubo A., Kawasaki H., Ebihara T., Hata K., Tanahashi S., Ichinose S., Imoto I., Inazawa J., Kudoh J., Amagai M. 2011. SASPase regulates stratum corneum hydration through pro filaggrin-to-filaggrin processing. EMBO Molecular Medicine, 3, 6: 320-333
- Mayr E., 1960. The emergence of evolutionary novelties. V: Evolution after Darwin. Tax S. (ed.). Chicago, University of Chicago Press: 349–380
- Meindl-Beinker, N. M., Matsuzaki, K., & Dooley, S. 2012. TGF-β Signaling in Onset and Progression of Hepatocellular Carcinoma. Digestive Diseases, 30, 5: 514-523
- Mi S., Lee X., Li X., Veldman G.M., Finnerty H., Racie L., LaVallie E., Tang X.Y., Edouard P., Howes S., Keith J.C. Jr, McCoy J.M. 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature, 403, 6771: 785–789
- Miller W.J., McDonald J.F., Nouaud D., Anxolabehere D. 1999. Molecular domestication – more than a sporadic episode in evolution. Genetica, 107: 197-207 doi:10.1007/978-94-011-4156-7\_22
- Miller M.A., Pfeiffer W., Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. V: Proceedings of the Gateway Computing Environments. Workshop (GCE), 14. november 2010. New Orleans, Association for Computing Machinery: 1-8
- Moczek A.P., 2008. On the origins of novelty in development and evolution. Bioessays, 30, 5: 432-447
- Muffato M., Louis A., Poisnel C.E., Roest Crollius H. 2010. Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. Bioinformatics, 26, 8: 1119-1121

- Musser A.M. 2003. Review of the monotreme fossil record and comparison of palaeontological and molecular data. Comparative Biochemistry and Physiology-Part A: Molecular & Integrative Physiology, 136, 4: 927-942
- Müller G.B. 1990. Developmental mechanisms at the origin of morphological novelty: A side-effect hypothesis. V: Evolutionary Innovations. Nitecki M.H. (ed.). Chicago, University of Chicago Press: 99–130
- Müller G.B., Newman S.A. 2005. The innovation triad: An EvoDevo agenda. Journal of Experimental Zoology (MDE), 304, 6: 487-503
- Nagasaki K., Manabe T., Hanzawa H., Maass N., Tsukada T., Yamaguchi K. 1999. Identification of a novel gene, LDOC1, down-regulated in cancer cell lines. Cancer Letters, 140, 1-2: 227-234
- Nagasaki K., Schem C., von Kaisenberg C., Biallek M., Rösel F., Jonat W., Maass N. 2003. Leucine-zipper protein, LDOC1, inhibits NF-kappaB activation and sensitizes pancreatic cancer cells to apoptosis. International Journal of Cancer, 105, 4: 454-458
- Neme R., Tautz D. 2013. Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. BMC Genomics, 14, 1: 117 doi:10.1186/1471-2164-14-117
- Nishihara H., Maruyama S., Okada N. 2009. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals.
  Proceedings of the National Academy of Sciences of the United States of America, 106, 13: 5235-5240
- Ohno S. 1984. Birth of a unique enzyme from an alternative reading frame of the preexisted, internally repetitious coding sequence. Proceedings of the National Academy of Sciences of the United States of America, 81, 8: 2421–2425
- Okabe H., Satoh S., Furukawa Y., Kato T., Hasegawa S., Nakajima Y., Yamaoka Y., Nakamura Y. 2003. Involvement of PEG10 in human hepatocellular carcinogenesis through interaction with SIAH1. Cancer Research, 63, 12: 3043-3048
- Okamura K., Nakai K., 2008. Retrotransposition as a source of new promoters. Molecular Biology and Evolution, 25, 6: 1231–1238
- Oliver K.R, Greene W.K. 2009. Transposable elements: powerful facilitators of evolution. Bioessays, 31, 7:703-714

- Oldham M.C., Horvath S., Geschwind D.H. 2006. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. Proceedings of the National Academy of Sciences of the United States of America, 103, 47: 17973-17978
- Ono R., Kobayashi S., Wagatsuma H., Aisaka K., Kohda T., Kaneko-Ishino T., Ishino F. 2001. A retrotransposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21. Genomics, 73, 2: 232-237
- Ono R., Nakamura K., Inoue K., Naruse M., Usami T., Wakisaka-Saito N., Hino T., Suzuki-Migishima R., Ogonuki N., Miki H., Kohda T., Ogura A., Yokoyama M., Kaneko-Ishino T., Ishino F. 2006. Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. Nature Genetetics, 38, 1: 101-106
- Ono R., Kuroki Y., Naruse M., Ishii M., Iwasaki S., Toyoda A., Fujiyama A., Shaw G., Renfree M.B., Kaneko-Ishino T., Ishino F. 2011. Identification of tammar wallaby SIRH12, derived from a marsupial-specific retrotransposition event. DNA Research, 18, 4: 211-219
- Parisi F., Sonderegger B., Wirapati P., Delorenzi M., Naef F. 2009. Relationship between estrogen receptor alpha location and gene induction reveals the importance of downstream sites and cofactors. BMC Genomics, 10,1: 381 doi: 10.1186/1471-2164-10-381
- Parsch J., Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. Nature Reviews Genetics, 14, 2: 83-87
- Piskurek O., Nishihara H., Okada N. 2009. The evolution of two partner LINE/SINE families and a full-length chromodomain-containing Ty3/Gypsy LTR element in the first reptilian genome of Anolis carolinensis. Gene, 441, 1: 111-118
- Plath N., Ohana O., Dammermann B., Errington M.L., Schmitz D., Gross C., Mao X., Engelsberg A., Mahlke C., Welzl H., Kobalz U., Stawrakakis A., Fernandez E., Waltereit R., Bick-Sander A., Therstappen E., Cooke S.F., Blanquet V., Wurst W., Salmen B., Bösl M.R., Lipp H.P., Grant S.G., Bliss T.V., Wolfer D.P., Kuhl D. 2006. Arc/Arg3.1 is essential for the consolidation of synaptic plasticity and memories. Neuron, 52, 3: 437-444
- Potrzebowski L., Vinckenbosch N., Kaessmann H. 2010) The emergence of new genes on the young therian X. Trends in Genetics, 26, 1: 1-4

- Preuss T.M., Cáceres M., Oldham M.C., Geschwind D.H. 2004. Human brain evolution: insights from microarrays. Nature Reviews Genetics, 5, 11: 850-860
- Reik W., Walter J. 2001. Genomic imprinting: parental influence on the genome. Nature Reviews Genetics, 2, 1: 21-32
- Ronquist F., Huelsenbeck J.P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 19, 12: 1572-1574
- Quentin Y. 1988. The Alu family developed through successive waves of fixation closely connected with primate lineage history. Journal of Molecular Evolution, 27, 3: 194-202
- Sabath N., Wagner A., Karlin D. 2012. Evolution of viral proteins originated de novo by overprinting. Molecular Biology and Evolution, 29, 12: 3767-3780
- Saitou N., Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4, 4: 406-425
- Sander T.L., Stringer K.F., Maki J.L., Szauter P., Stone J.R., Collins T. 2003. The SCAN domain defines a large family of zinc finger transcription factors. Gene, 310: 29-38
- Seitz H., Youngson N., Lin S.P., Dalbert S., Paulsen M., Bachellerie J.P., Ferguson-Smith A.C., Cavaillé J., 2003. Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. Nature Genetics, 34, 3: 261-262
- Sekita Y., Wagatsuma H., Nakamura K., Ono R., Kagami M., Wakisaka N., Hino T., Suzuki-Migishima R., Kohda T., Ogura A., Ogata T., Yokoyama M., Kaneko-Ishino T., Ishino F. 2008. Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. Nature Genetics. 40, 2: 243-248
- Schmitz J., Brosius J. 2011. Exonization of transposed elements: A challenge and opportunity for evolution. Biochimie, 93, 11: 1928-1934
- Schüller M., Jenne D., Voltz R., 2005. The human PNMA family: novel neuronal proteins implicated in paraneoplastic neurological disease. Journal of Neuroimmunology, 169, 1: 172-176
- Schwartz A., Chan D.C., Brown L.G., Alagappan R., Pettay D., Disteche C., McGillivray B., de la Chapelle A., Page D.C. 1998. Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by

Yp inversion through LINE-LINE recombination. Human Molecular Genetics, 7, 1: 1-11

- Shedlock A.M., Edwards S.V. 2009. Amniotes (Amniota). V: The Timetree of Life,. Hedges S. B., Kumar S. (eds.). New York. Oxford University Press: 375-379
- Singer S. S., Maennel D. N., Hehlgans T., Brosius J., Schmitz J. 2004. From "junk" to gene: curriculum vitae of a primate receptor isoform gene. Journal of Molecular Biology, 341, 4: 883-886
- Sinzelle L., Izsvák Z., Ivics Z. 2009. Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cellular and Molecular Life Sciences, 66, 6: 1073-1093
- Sorek R., Ast G., Graur D., 2002. Alu-containing exons are alternatively spliced. Genome Research, 12, 7: 1060-1067
- Sorek R. 2007. The birth of new exons: Mechanisms and evolutionary consequences. RNA, 13, 10: 1603-1608
- Stark A., Brennecke J., Bushati N., Russell R.B., Cohen S.M., 2005. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell, 123, 6: 1133-1146
- Suzuki S., Renfree M.B., Pask A.J., Shaw G., Kobayashi S., Kohda T., Kaneko-Ishino T., Ishino F. 2005. Genomic imprinting of IGF2, p57(KIP2) and PEG1/MEST in a marsupial, the tammar wallaby. Mechanisms of Development, 122, 2: 213-222
- Suzuki S., Ono R., Narita T., Pask A.J., Shaw G., Wang C., Kohda T., Alsop A.E., Marshall Graves J.A., Kohara Y., Ishino F., Renfree M.B., Kaneko-Ishino T. 2007. Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. PLoS Genetics, 3, 4: e55 doi:10.1371/journal.pgen.0030055
- Takaji M., Komatsu Y., Watakabe A., Hashikawa T., Yamamori T. 2009. Paraneoplastic antigen-like 5 gene (PNMA5) is preferentially expressed in the association areas in a primate specific manner. Cerebral Cortex, 19, 12: 2865-2879
- Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution, 28, 10: 2731-2739

- Tan K.O., Tan K.M., Chan S.L., Yee K.S., Bevort M., Ang K.C., Yu V.C. 2001. MAP-1, a novel proapoptotic protein containing a BH3-like motif that associates with Bax through its Bcl-2 homology domains. Journal of Biological Chemistry, 276, 4: 2802-2807
- Tang Y., Katuri V., Dillner A., Mishra B., Deng C. X., Mishra L. 2003. Disruption of transforming growth factor-β signaling in ELF β-spectrin-deficient mice. Science, 299, 5606: 574-577
- Tse W.T., Tang J., Jin O., Korsgren C., John K.M., Kung A.L., Gwynn B., Peters L.L., Lux S.E. 2001. A new spectrin, beta IV, has a major truncated isoform that associates with promyelocytic leukemia protein nuclear bodies and the nuclear matrix. Journal of Biological Chemistry, 276, 26: 23974-23985
- Ullu E., Tschudi C. 1984. Alu sequences are processed 7SL RNA genes. Nature, 312, 5990: 171-172
- Van de Peer Y., Maere S., Meyer A. 2009. The evolutionary significance of ancient genome duplications. Nature Reviews Genetics, 10, 10: 725–732
- Volff J.N., Körting C., Altschmied J., Duschl J., Sweeney K., Wichert K., Froschauer A., Schartl M. 2001. Jule from the fish Xiphophorus is the first complete vertebrate Ty3/Gypsy retrotransposon from the Mag family. Molecular Biology and Evolution, 18, 2: 101–111
- Volff J.N., Bouneau L., Ozouf-Costaz C., Fischer C. 2003. Diversity of retrotransposable elements in compact pufferfish genomes. Trends in Genetics, 19, 12: 674-678
- Volff J.N. 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays, 28, 9: 913-922
- Walsh C. P., Chaillet J. R., Bestor T. H. 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nature Genetics, 20, 2: 116
- Wang P. J., McCarrey J. R., Yang F., Page D. C. 2001. An abundance of X-linked genes expressed in spermatogonia. Nature Genetics, 27, 4: 422-426
- Wicker T., Sabot F., Hua-Van A., Bennetzen J.L., Capy P., Chalhoub B., Flavell A., Leroy P., Morgante M., Panaud O., Paux E., SanMiguel P., Schulman A.H. 2007. A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics, 8, 12: 978-982

- Wills N.M., Moore B., Hammer A., Gesteland R.F., Atkins J.F. 2006. A functional -1 ribosomal frameshift signal in the human paraneoplastic Ma3 gene. Journal of Biological Chemistry, 281, 11: 7082-7088
- Winkelmann J. C., Chang J. G., Tse W. T., Scarpa A. L., Marchesi V. T., Forget B. G. 1990. Full-length sequence of the cDNA for human erythroid betaspectrin. Journal of Biological Chemistry, 265, 20: 11827-11832
- Wu X., Schmidt J.A., Avarbock M.R., Tobias J.W., Carlson C.A., Kolon T.F., Ginsberg J.P., Brinster R.L. 2009. Prepubertal human spermatogonia and mouse gonocytes share conserved gene expression of germline stem cell regulatory molecules. Proceedings of the National Academy of Sciences of the United States of America, 106, 51: 21672-21677
- Yamamori T. 2011. Selective gene expression in regions of primate neocortex: implications for cortical specialization. Progress in Neurobiology, 94, 3: 201–222
- Yoder J.A., Walsh, C.P., Bestor, T.H. 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends in Genetics, 13, 8: 335-340
- Yuan Y.W., Wessler S.R. 2011. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. Proceedings of the National Academy of Sciences of the United States of America, 108, 19: 7884-7889
- Zarnack K., König J., Tajnik M., Martincorena I., Eustermann S., Stévant I., Reyes A., Anders S., Luscombe N.M., Ule J. 2013. Direct Competition between hnRNP C and U2AF65 Protects the Transcriptome from the Exonization of Alu Elements. Cell, 152, 3: 453-466
- Zdobnov E.M., Campillos M., Harrington E.D., Torrents D., Bork P. 2005. Protein coding potential of retroviruses and other transposable elements in vertebrate genomes. Nucleic Acids Research, 33, 3: 946-954

# ZAHVALA

Doktorsko delo sem opravljal na Odseku za molekularne in biomedicinske znanosti Instituta "Jožef Stefan" v Ljubljani.

Iskreno se zahvaljujem mentorju doc. dr. Dušanu Kordišu za vse nasvete, spodbude in uspešno vodenje pri pripravi doktorskega dela.

Zahvaljujem se vsem sodelavcem in prijateljem na Institutu, ki so mi velikokrat polepšali dan in nudili pomoč, ko je to bilo potrebno.

Zahvaljujem se tudi moji družini za vso podporo in ljubezen.

# PRILOGA A

# Preglednica P1: Seznam genov uporabljenih pri filogenomskih analizah

Gen	Organizem	Taksonomska skupina	Številka zaporedja	Kromosomska lokacija
PEG10				
	Ailuropoda melanoleuca	Boreoeutheria	ACTA01043706.1	
	Bos taurus	Boreoeutheria	NP_001165989.1	
	Callithrix jacchus	Boreoeutheria	ACFV01062247.1	
	Canis lupus familiaris	Boreoeutheria	NP_001166014	
	Cavia porcellus	Boreoeutheria	AAKN02016859.1	
	Choloepus hoffmanni	Xenarthra	ABVD01060229	
	Dasypus novemcinctus	Xenarthra	AAGV03171820.1	
	Dipodomys ordii	Boreoeutheria	ABRO01072656.1	
	Echinops telfairi	Afrotheria	AAIY01557043	
	Equus caballus	Boreoeutheria	NP_001166029	
	Felis catus	Boreoeutheria	AANG02100824.1	
	Gorilla gorilla	Boreoeutheria	CABD02345544.1	
	Heterocephalus glaber	Boreoeutheria	AFSB01016640.1	
	Homo sapiens	Boreoeutheria	NP_055883	7 (7q21)
	Ictidomys tridecemlineatus	Boreoeutheria	AGTP01068175.1	
	Loxodonta africana	Afrotheria	AAGU03015863.1	
	Macaca mulatta	Boreoeutheria	NP_001165893.1	
	Macropus eugenii	Marsupialia	 ABQ0020785483.1	
	Microcebus murinus	Boreoeutheria	ABDC01270285.1	
	Mus musculus	Boreoeutheria	NP 570947	
	Mustela putorius furo	Boreoeutheria	AEYP01068702.1	
	Mvotis lucifugus	Boreoeutheria	AAPE02008825.1	
	Nomascus leucogenvs	Boreoeutheria	ADFV01000933.1	
	Orvetolagus cuniculus	Boreoeutheria	AAGW02035426.1	
	Otolemur garnettii	Boreoeutheria	AAOR03012733.1	
	Pan troglodytes	Boreoeutheria	AACZ03052811.1	
	Pongo abelii	Boreoeutheria	ABGA01007342.1	
	Procavia capensis	Afrotheria	ABR001517954.1	
	Pteropus vampyrus	Boreceutheria	ABRP01146153.1	
	Sarcophilus harrisii	Marsunialia	AFFY01088593 1	
	Sarcophilus harristi Sorex graneus	Boreceutheria	A AI T01650199 1	
	Sus scrofa	Boreceutheria	NP_001120696	
	Trichochus manatus latirostris	Afrotheria	AHIN01116204 1	
	Tursions truncatus	Boreceutheria	ABRN02/00/00 1	
CACA	Turstops trancatus	Doreocutienta	ABR(1024)040).1	
0A04	Ros taurus	Borecoutheria	A AEC03124804	
	Callithrix jacchus	Boreoeutheria	ACEV01168828 1	
	Canis lupus familiaris	Boreoeutheria	AAEX03026494	
	Cavia porcellus	Boreoeutheria	AAKN02029999	
	Dasypus novemcinctus	Xenarthra	AAGV03221368.1	
	Dipodomys ordii	Boreoeutheria	ABRO01238289.1	
	Equus caballus	Boreoeutheria	AAWR02039660.1	
	Felis catus	Boreoeutheria	ACBE01497584.1	
	Gorilla gorilla	Boreoeutheria	CABD02400541.1	
	Heterocephalus glaber	Boreoeutheria	AFSB01128218.1	V (V 10 1)
	Homo sapiens	A frotheria	NP_001019626.1	X (Xq13.1)
	Loxoaonia africana Macaca mulatta	Boreceutheric	AAGUU3048013 FHH30837	
	Microcebus murinus	Boreceutheria	ABDC01193954-1	
	Mus musculus	Boreoeutheria	NP 899141	
		Doreocumeria		an madalinia

nadarjevanje	pregledinee 11. Sezhan genov			
Gen	Organizem	Taksonomska	Stevilka zaporedja	Kromosomska
		skupina		lokacija
	Mustela putorius furo	Boreoeutheria	AES05561.1	
	Myotis lucifugus	Boreoeutheria	AAPE02004103	
	Nomascus leucogenvs	Boreoeutheria	ADFV01127123.1	
	Ochotona princens	Boreceutheria	AAYZ01235013.1	
	Orvetolagus cuniculus	Boreoeutheria	AAGW02057529	
	Oryctolagus cuniculus	Boreceutheria	A AGW02057529 1	
	Otyciolagus cunicalas	Borecoutheria	AAG W02057525.1	
	Dan troaledutes	Borecoutheria	AAQK03183990	
	Pan inogloayles	Doreoeuuleria Dana a suth ania	AAC205159464.1	
	Pongo abelli	Boreoeutneria	ABGA0132/151.1	
	Procavia capensis	Airotneria	ABRQ01198437.1	
	Rattus norvegicus	Boreoeutheria	AABR05180029.1	
	Sorex araneus	Boreoeutheria	AAL101549896.1	
	Trichechus manatus latirostris	Afrotheria	AHIN01073765.1	
	Tupaia belangeri	Boreoeutheria	AAPY01750682.1	
	Tursiops truncatus	Boreoeutheria	ABRN02257176.1	
RGAG1				
	Bos taurus	Boreoeutheria	XP_001788607	
	Canis lupus familiaris	Boreoeutheria	AAEX03026772.1	
	Cavia porcellus	Boreoeutheria	AAKN02039973.1	
	Dasypus novemcinctus	Xenarthra	AAGV03179636.1	
	Eauus caballus	Boreceutheria	AAWR02039240 1	
	Equis cabanas Folis catus	Boreceutheria	ACBE01502664 1	
	Gorilla gorilla	Boreceutheria	CABD02405307 1	
	Hotoroconhalus alabor	Borecoutheria	AHKG01071015 1	
	Homo sapions	Borecoutheria	ND 065820 1	$\mathbf{V}(\mathbf{V}_{2}22)$
	Internet and the second s	Borecoutheria	ACTD01020640 1	A (A423)
	Long donta africana	Afrethania	AGTF01029049.1	
	Loxodonia africana Managana andatta	Alfouneria Deve e costie	AAGUU5050054.1	
	Macaca mulalla	Doreoeutieria Dana a suth ania	NP_001025524.1	
	Mus musculus	Boreoeutneria	NP_001035524.1	
	Mustela putorius furo	Boreoeutheria	AGIQ01035711.1	
	Myotis lucifugus	Boreoeutheria	AAPE02002/54.1	
	Nomascus leucogenys	Boreoeutheria	ADFV01008253.1	
	Ochotona princeps	Boreoeutheria	AAYZ01396442.1	
	Oryctolagus cuniculus	Boreoeutheria	AAGW02041334.1	
	Otolemur garnettii	Boreoeutheria	AAQR03155329.1	
	Pan troglodytes	Boreoeutheria	AACZ03136290.1	
	Pongo abelii	Boreoeutheria	ABGA01005524.1	
	Pteropus vampyrus	Boreoeutheria	ABRP01260471.1	
	Rattus norvegicus	Boreoeutheria	AAHX01113452.1	
	Tarsius syrichta	Boreoeutheria	ABRT010081689.1	
	Trichechus manatus latirostris	Afrotheria	AHIN01071955.1	
	Tupaia belangeri	Boreoeutheria	AAPY01382729.1	
RTL1	7			
	Bos taurus	Boreceutheria	NP 0011810/0	
	Callithrix jacohus	Borecoutheria	ACEV01132248 1	
	Cania lunus familiaria	Borecoutheria	ACT V01132240.1	
	Canis inpus juminaris	Doreoeuuleria Dana a suth ania	AAEX03000004	
	Cavia porcellus	Boreoeutneria Vananthua	AAKN02051948.1	
	Choioepus noffmanni	Xenarthra	ABVD01521525.1	
	Dasypus novemcinctus	Aenarthra	AAGVU3256/35.1	
	Daubentonia madagascariensis	Boreoeutheria	AG1M011/56538.1	
	Dipodomys ordii	Boreoeutheria	ABRO01195543.1	
	Equus caballus	Boreoeutheria	AAWR02020062.1	
	Felis catus	Boreoeutheria	AANG02058241.1	
	Gorilla gorilla	Boreoeutheria	CABD02116350.1	
	Heterocephalus glaber	Boreoeutheria	EHB11169.1	
	Homo sapiens	Boreoeutheria	AAI50618	14 (14q32.2)
	Ictidomys tridecemlineatus	Boreoeutheria	AGTP01074072.1	
				se nadaljuje

Con	Organizam	Taksonomska	Štovilka zanorodia	Kromosomska
Gen	Organizem	r aksonomska skuping	Steviika zaporeuja	lokacija
	Loxodonta africana	Afrotheria	A AGU03025978	юкастја
	Macaca mulatta	Boreceutheria	XP 001110319	
	Microcebus murinus	Boreoeutheria	ABDC01565557.1	
	Mus musculus	Boreoeutheria	NP 908998	
	Mustela putorius furo	Boreoeutheria	AEYP01037657.1	
	Myotis lucifugus	Boreoeutheria	AAPE02047295.1	
	Nomascus leucogenvs	Boreoeutheria	ADFV01092893.1	
	Ochotona princeps	Boreoeutheria	AAYZ01391477	
	Oryctolagus cuniculus	Boreoeutheria	AAGW02077263.1	
	Otolemur garnettii	Boreoeutheria	AAQR03046506.1	
	Pan troglodytes	Boreoeutheria	AADA01312366.1	
	Pongo abelii	Boreoeutheria	ABGA01039457.1	
	Procavia capensis	Afrotheria	ABRQ01643018	
	Pteropus vampyrus	Boreoeutheria	ABRP01223356.1	
	Trichechus manatus latirostris	Afrotheria	AHIN01054610.1	
	Tursiops truncatus	Boreoeutheria	ABRN02353414.1	
ZCCHC16				
	Callithrix jacchus	Boreoeutheria	ACFV01055746.1	
	Canis lupus familiaris	Boreoeutheria	XP_549194	
	Choloepus hoffmanni	Xenarthra	ABVD01449713.1	
	Daubentonia madagascariensis	Boreoeutheria	AGTM011637388.1	
	Echinops telfairi	Afrotheria	AAIY01429107.1	
	Equus caballus	Boreoeutheria	XP_001488764	
	Felis catus	Boreoeutheria	AANG02074530.1	
	Gorilla gorilla	Boreoeutheria	CABD02405487.1	
	Homo sapiens	Boreoeutheria	NP_001004308	X (Xq23)
	Ictidomys tridecemlineatus	Boreoeutheria	AAQQ01371003.1	
	Loxodonta africana	Afrotheria	AAGU03055959.1	
	Macaca mulatta	Boreoeutheria	NP_001181102	
	Mus musculus	Boreoeutheria	NP_001028967	
	Mustela putorius furo	Boreoeutheria	AEYP01071779.1	
	Myotis lucifugus	Boreoeutheria	AAPE02002693	
	Nomascus leucogenys	Boreoeutheria	ADFV01008333.1	
	Oryctolagus cuniculus	Boreoeutheria	XP_002720260	
	Otolemur garnettii	Boreoeutheria	AAQR03170339.1	
	Pan troglodytes	Boreoeutheria	AADA01038094.1	
	Procavia capensis	Afrotheria	ABRQ01131885.1	
	Trichechus manatus latirostris	Afrotheria	AHIN01072081.1	
ZCCHC5				
	Equus caballus	Boreoeutheria	XP_003365869.1	
	Gorilla gorilla	Boreoeutheria	CABD02401497.1	
	Homo sapiens	Boreoeutheria	AAI36549	X (Xq21.1)
	Ictidomys tridecemlineatus	Boreoeutheria	AGTP01054468.1	
	Macaca mulatta	Boreoeutheria	AANU01174818.1	
	Mus musculus	Boreoeutheria	NP_955/62	
	Myotis lucifugus	Boreoeutneria	AAPE02003884.1	
	Nomascus leucogenys	Boreoeutheria	ADF V01047424.1	
	Oryctolagus cuniculus	Boreoeutheria	AAGW02041680.1	
	r un trogiodytes Pongo abalii	Doreoeutheria	AADAU1082408.1	
	r ongo abelli Ptaropus vampurus	Boreceutheria	ADUAU1280131.1 ENSDVA D0000002872	
	r teropus vampyrus	Borecoutheric	LINSEVAEUUUUUUUUU00/2 ARDD01216211-1	
	r ieropus vampyrus	Afrothania	ADKPU1210211.1	
	1 ricnecnus manatus latirostris	Airotneria	AHIINU100000/.1	
	Tursions truncatus	Borecoutheric	AAF I UI / JU949 ABDN02156766 1	
IDOCI	raisiops irancatas	Doreoeuuleria	ABIN02130700.1	
	Pos taunus	Doracouthania	ND 001075012 1	
	Dos laurus Callithrir igoobus	Borecoutheric	MP_001073013.1 ACEV01165274-1	
	Cummux jacentus	Boreoeutileria	ACF V01103274.1	sa nadalinia
				se naualjuje

Gen	Organizem	Taksonomska	Številka zaporedja	Kromosomsk
		skupina		lokacija
	Canis lupus familiaris	Boreoeutheria	AAEX03026983.1	
	Cavia porcellus	Boreoeutheria	XP_003464753.1	
	Dasypus novemcinctus	Xenarthra	AAGV03124563.1	
	Dipodomys ordii	Boreoeutheria	ABRO01183309.1	
	Echinops telfairi	Afrotheria	AAIY01054437	
	Equus caballus	Boreoeutheria	AAWR02041332.1	
	Erinaceus europaeus	Boreoeutheria	AANN01096519.1	
	Felis catus	Boreoeutheria	AANG02125169.1	
	Gorilla gorilla	Boreoeutheria	CABD02409225.1	
	Heterocephalus glaber	Boreoeutheria	AHKG01074095.1	
	Homo sapiens	Boreoeutheria	NP_036449.1	X (Xq27)
	Ictidomys tridecemlineatus	Boreoeutheria	AGTP01037783.1	
	Macaca mulatta	Boreoeutheria	XP_001085397.1	
	Mus musculus	Boreoeutheria	NP_001018097.1	
	Mustela putorius furo	Boreoeutheria	AEYP01102792.1	
	Myotis lucifugus	Boreoeutheria	AAPE02042786.1	
	Nomascus leucogenys	Boreoeutheria	ADFV01118802.1	
	Ochotona princeps	Boreoeutheria	AAYZ01654760.1	
	Oryctolagus cuniculus	Boreoeutheria	AAGW02055714.1	
	Otolemur garnettii	Boreoeutheria	AAQR03136356.1	
	Pan troglodytes	Boreoeutheria	AACZ03140206.1	
	Pteropus vampyrus	Boreoeutheria	ABRP01249445	
	Trichechus manatus latirostris	Afrotheria	AHIN01080024.1	
	Tupaia belangeri	Boreoeutheria	AAPY01387257.1	
	Tursiops truncatus	Boreoeutheria	ABRN02225857.1	
DOCIL	I			
DUCIL	Bos taurus	Boreceutheria	XP 0012538901	
	Callithrix jacchus	Boreceutheria	ACEV01163893 1	
	Canis lupus familiaris	Boreceutheria	XP 538334 1	
	Cavia norcellus	Boreceutheria	A AKN02056344	
	Dasynus novemeinetus	Venarthra	A AGV03215739 1	
	Eaus caballus	Roreceutheria	AAWR02040740 1	
	Equis cabus	Boreceutheria	A ANG02167187 1	
	Heterocenhalus alaber	Boreceutheria	AHKC01032576 1	
	Homo senions	Doreceutheria	ND 115662.2	22(22a12,21)
	Internet descentine atus	Boreceutheria	AGTD01008805 1	22 (22415.51)
	Lorodonta africana	Afrotheria	AGTF01098895.1 A AGU03077096 1	
	Loxodonia ajricana Magaga mulatta	Anotheria Domocoutheria	ND 001181272 1	
	Macaca mulalla	Doreoeutheria	NP_001181372.1	
	Mustala mutaning funa	Boreceutheria	NP_808298.2	
	Mustela putorius furo	Boreoeutheria	AGIQ01046255.1	
	Myotis lucifugus Muotia lucifugus	Boreceutheria	AAPE02062099	
	Myotis iucifugus	Boreoeutheria	AAPE02062099.1	
	Nomascus leucogenys	Boreoeutheria	ADF V011320/8.1	
	Otolemur garnettu	Boreoeutheria	AAQR03015785.1	
	Ovis aries	Boreoeutheria	ACIV010481667.1	
	Pan troglodytes	Boreoeutheria	AADA01102/19.1	
	Pongo abelii	Boreoeutheria	ABGA01382764.1	
	Procavia capensis	Afrotheria	ABRQ01229456.1	
	Pteropus vampyrus	Boreoeutheria	ABKP01299088.1	
	Kattus norvegicus	Boreoeutheria	AAHX01051963.1	
	Tarsius syrichta	Boreoeutheria	ABR1010/80528.1	
	Trichechus manatus latirostris	Atrotheria	AHIN01072820.1	
	Tupaia belangeri	Boreoeutheria	AAPY01023583.1	
	Tursiops truncatus	Boreoeutheria	ABRN02316718.1	
FAM127				
	Bos taurus	Boreoeutheria	NP_001068990	
	Callithrix jacchus	Boreoeutheria	ACFV01112550.1	
	Canis lupus familiaris	Boreoeutheria	XP_003435578	
				se nadaliuje

Gen	Organizem	Taksonomska	Številka zaporedja	Kromosomska lokacije
	Dagunus pouroinatus	Vanarthro	A A CW02227800 1	iunacija
	Dinodomus ordii	Aenaruna Porocouthorio	AAG V 05257809.1	
	Dipodomys ordii Echinope tolfairi	Afrothoria	ADKOU1105821.1 A AIV01062207 1	
	Echinops leijain Echinops leijain	Anomena Doracouthoria	AAI101003307.1	
	Equus caballus Ening come come come	Doreceutheria	AAWK02054005.1	
	Corilla corilla	Boreceutheria	CADD02408507 1	
	Gorilla gorilla	Doreceutheria	CADD02406307.1	
	Hererocephalus glaber	Doreceutheria	AFSD01150167.1	$\mathbf{V}(\mathbf{V}_{\alpha}\mathbf{M})$
	Homo sapiens	Boreceutheria	NP_0010/1059.1	A (Aq20)
	Lorodonta africana	Afrothoria	AGIF01094707.1	
	Loxodonia africana Macaca mulatta	Anotheria Device south suite	ND 001181065	
	Micaca mutatta Micaco chua munimus	Boreceutheria	NP_001181005	
	Microcebus murinus	Doreoeutheria	ABDC01554599.1	
	Must al a sect a size from	Doreoeutheria	AAH 101/45100.1	
	Mustela putorius juro Mustia lusifusua	Boreceutheria	AGIQ01047722.1 A ADE02002176 1	
	Myous lucijugus	Doreoeutheria	AAPE02002170.1	
	Nomascus leucogenys	Boreceutheria	ADF V010/2099.1	
	Omotolagua auriculua	Boreceutheria	AAIZU14918// AACW020420101	
	Oryctolagus cuniculus	Boreoeutheria	AAGWU2U40818.1	
	Otolemur garnettii	Boreoeutheria	AAQR03189099	
	Otolemur garnettii	Boreoeutheria	AAQK03189099.1	
	Pan troglodytes	Boreoeutheria	AADA01342505.1	
	Pongo abelli	Boreoeutneria	ABGA01295535.1	
	Procavia capensis	Airotheria	ABRQ01574575.1	
	Pteropus vampyrus	Boreoeutheria	ABRP01194706.1	
	Tarsius syrichta	Boreoeutheria	ABR1011084406.1	
	Turstops truncatus	Boreoeutheria	ABRN02215744.1	
CIDII 12	vicugna pacos	Boreoeutneria	ABRR01323960.1	
SIKH12	Macronus auganii	Marcunialia	ABOO020446584 1	
	macropus eugenii	Warsuplana	ADQ0020440384.1	
C22orf29				
-	Callithrix jacchus	Boreoeutheria	XP_002743589.1	
	Dipodomys ordii	Boreoeutheria	ABRO01388190.1	
	Echinops telfairi	Afrotheria	ENSETET00000016288	
	Homo sapiens	Boreoeutheria	NP_078903.3	22 (22q11.21)
	Ictidomys tridecemlineatus	Boreoeutheria	AGTP01078122.1	
	Loxodonta africana	Afrotheria	AAGU03075386.1	
	Macaca mulatta	Boreoeutheria	AFE75991.1	
	Nomascus leucogenys	Boreoeutheria	XP_003280408.1	
	Oryctolagus cuniculus	Boreoeutheria	AAGW02072302.1	
	Pan troglodytes	Boreoeutheria	NP_001230896.1	
	Pongo abelii	Boreoeutheria	XP_002830899.1	
	Trichechus manatus latirostris	Afrotheria	AHIN01144892.1	
PNMA1				
	Ailuropoda melanoleuca	Boreoeutheria	XP_002914729	
	Bos taurus	Boreoeutheria	NP_001098461	
	Callithrix jacchus	Boreoeutheria	XP_002754140	
	Canis lupus familiaris	Boreoeutheria	XP_547896	
	Cavia porcellus	Boreoeutheria	XP_003472455.1	
	Choloepus hoffmanni	Xenarthra	ABVD01858941	
	Cricetulus griseus	Boreoeutheria	XP_003505180.1	
	Dasypus novemcinctus	Xenarthra	AAGV03070378.1	
	Echinops telfairi	Afrotheria	AAIY01462143.1	
	Equus caballus	Boreoeutheria	XP 001490004	
	Felis catus	Boreoeutheria	ACBE01211376	
	Gorilla gorilla	Boreceutheria	CABD02112671	
	Heterocephalus olaber	Boreceutheria	AFSB01202249 1	
	Surrium Surreit	2 or e o e a uner i u		
	Homo sapiens	Boreoeutheria	NP 006020	14 (14a24.3)

Com	Organizam	Tobaonomako	Štavilka ganaradia	Vromocomelro
Gen	Organizem	Такзопотска	Steviika zaporedja	Kromosomska
	~	skupina		lokacija
	Ictidomys tridecemlineatus	Boreoeutheria	AGTP01020471.1	
	Loxodonta africana	Afrotheria	XP_003408802.1	
	Macaca mulatta	Boreoeutheria	XP_001091302	
	Mus musculus	Boreoeutheria	NP_081714	
	Mustela putorius furo	Boreoeutheria	AGTQ01032725	
	Myotis lucifugus	Boreoeutheria	AAPE02034650	
	Nomascus leucogenys	Boreoeutheria	XR_121367.1	
	Oryctolagus cuniculus	Boreoeutheria	XP_002719706.1	
	Otolemur garnettii	Boreoeutheria	AAQR03045035	
	Pan troglodytes	Boreoeutheria	XP_001141691	
	Pongo abelu	Boreoeutheria	XP_002824970	
	Pteropus vampyrus	Boreoeutneria	ABRP01196995	
	Ratius norvegicus	Boreoeutneria	NP_570833	
	Sorex araneus	Boreoeutneria	AAL101503131	
	Sus scroja Tansius purichta	Boreoeutneria	AP_001928990	
	Tarsius syricnia	Boreoeutneria	ABR1010588175	
	Trichechus manatus latirostris	Airotheria	AHINU1016586.1	
	Tupala belangeri Tupala setua setua	Boreoeutheria	AAP 101355955	
PNMAD	Turstops truncatus	Doreoeutiierta	ABRIN01411903	
1 10101/12	Ailuropoda melanoleuca	Boreceutheria	ACTA01184811.1	
	Bos taurus	Boreceutheria	NP 001039938	
	Callithrix jacchus	Boreoeutheria	XP 002756910.1	
	Canis lupus familiaris	Boreceutheria	XP 543234	
	Choloepus hoffmanni	Xenarthra	ABVD01322223.1	
	Cricetulus griseus	Boreoeutheria	XP 003505211.1	
	Dasypus novemcinctus	Xenarthra	AAGV03058646.1	
	Daubentonia madagascariensis	Boreoeutheria	AGTM010363728.1	
	Echinops telfairi	Afrotheria	AAIY01539302	
	Equus caballus	Boreoeutheria	XP_001494457.1	
	Felis catus	Boreoeutheria	ACBE01547290.1	
	Gorilla gorilla	Boreoeutheria	CABD02358444.1	
	Heterocephalus glaber	Boreoeutheria	AFSB01189534	
	Homo sapiens	Boreoeutheria	NP_009188	8 (8p21.2)
	Ictidomys tridecemlineatus	Boreoeutheria	AGTP01103703.1	
	Loxodonta africana	Afrotheria	AAGU03044807	
	Macaca mulatta	Boreoeutheria	XP_001108453.1	
	Microcebus murinus	Boreoeutheria	ABDC01025881	
	Mus musculus	Boreoeutheria	NP_780707	
	Mustela putorius furo	Boreoeutheria	AEYP01085655	
	Myotis lucifugus	Boreoeutheria	AAPE02029833.1	
	Nomascus leucogenys	Boreoeutheria	XP_003272943.1	
	Oryctolagus cuniculus	Boreoeutheria	XP_002709354.1	
	Otolemur garnettii	Boreoeutheria	AAQR03106253.1	
	Pan troglodytes	Boreoeutheria	XP_528094	
	Pongo abelii	Boreoeutheria	NP_001127009	
	Procavia capensis	Afrotheria	ABRQ01301107.1	
	Rattus norvegicus	Boreoeutheria	NP_001100742	
	Sorex araneus	Boreoeutheria	AALT01151904	
	Sus scrofa	Boreoeutheria	XP_001925541.2	
	Tarsius syrichta	Boreoeutheria	ABR1010502723.1	
	<i>1 richechus manatus latirostris</i>	Afrotheria	AHINU1086405	
	Tupata belangeri	Boreoeutheria	AAPYU1194613.1	
ΜΟΛΡΙ	1 urstops truncatus	Богеоецинегна	ADKINU24193//.1	
	Callithrix jacchus	Boreceutheria	XP 0027542621	
	Cricetulus griseus	Boreceutheria	XP 003506983 1	
	0			

Gen	Organizem	Taksonomska	Številka zanoredia	Kromosomska
Gen	Organizeni	skunina	Steviika zaporeuja	lokacija
	Dasynus novemcinctus	Xenarthra	AAGV03218223 1	lokacija
	Daubentonia madagascariensis	Boreoeutheria	AGTM011692791.1	
	Felis catus	Boreoeutheria	ACBE01574473	
	Gorilla gorilla	Boreoeutheria	CABD02115479.1	
	Homo sapiens	Boreoeutheria	NP_071434	
	Ictidomys tridecemlineatus	Boreoeutheria	AAQQ01349199.1	
	Loxodonta africana	Afrotheria	XP_003408887.1	
	Macaca mulatta	Boreoeutheria	XP_001089203	
	Mus musculus	Boreoeutheria	NP_001136409	
	Mustela putorius furo	Boreoeutheria	AEYP01037879.1	
	Nomascus leucogenys	Boreoeutheria	XP_003260955.1	
	Otolemur garnettii	Boreoeutheria	AAQR03046215	
	Pan troglodytes	Boreoeutheria	XP_510137	
	Pongo abelii	Boreoeutheria	XP_002825093.1	
	Procavia capensis	Afrotheria	ABRQ01189423.1	
	Rattus norvegicus	Boreoeutheria	NP_001013119	
	Spermophilus tridecemlineatus	Boreoeutheria	AAQQ01349199	
D1/1-5/-5	Trichechus manatus latirostris	Atrotheria	AHIN01054316.1	
PNMA3		<b>—</b>		
	Callithrix jacchus	Boreoeutheria	XP_002763430.1	
	Dasypus novemcinctus	Xenarthra	AAGV03164000.1	
	Equus caballus	Boreoeutheria	XP_001496576.1	
	Felis catus	Boreoeutheria	AANG01462554	X (X 00)
	Homo sapiens	Boreoeutheria	NP_037496	X (Xq28)
	Ictidomys tridecemlineatus	Boreoeutheria	AGTP01098202.1	
	Loxodonta africana	Afrotheria	AAGU03085805.1	
	Macaca mulatta	Boreoeutheria	XP_001083094	
	Mus musculus	Boreoeutheria	NP_694809	
	Mustela putorius furo	Boreoeutheria	AGTQ01072216.1	
	Myotis lucifugus	Boreoeutheria	AAPE02015504	
	Nomascus leucogenys	Boreoeutheria	XP_003271923.1	
	Pan troglodytes	Boreoeutheria	XP_001145572.1	
	Pongo abelii	Boreoeutheria	XP_002832318.1	
	Pteropus vampyrus	Boreoeutheria	ABRP01300743.1	
	Rattus norvegicus	Boreoeutheria	NP_001099812	
PNMA5				
	Callithrix jacchus	Boreoeutheria	XP_002763433.1	
	Cavia porcellus	Boreoeutheria	AAKN02056074	
	Dasypus novemcinctus	Xenarthra	AAGV03164006.1	
	Dipodomys ordu	Boreoeutheria	ABK001400221.1	
	Equus caballus Corilla acrilla	Borecoutheria	AF_003303931.1 CARD02410420.1	
	Homo sapiens	Boreceutheria	VADD02410430.1 ND 001006620	$\mathbf{X}(\mathbf{X}_{0}28)$
	Itomo suprens	Boreceutheria	AGTD01090020	A (Aq20)
	Lorodonta africana	Afrotheria	XP 003/21289 1	
	Macaca mulatta	Boreceutheria	XP_001082427_1	
	Microcebus murinus	Boreoeutheria	ABDC01277479	
	Mus musculus	Boreceutheria	NP 001093931	
	Myotis lucifugus	Boreoeutheria	AAPE02015506.1	
	Nomascus leucogenys	Boreoeutheria	XP_003271921.1	
	Otolemur garnettii	Boreoeutheria	AAQR03167750	
	Pteropus vampyrus	Boreoeutheria	ABRP01159492	
	Rattus norvegicus	Boreoeutheria	NP_001101049	
	Trichechus manatus latirostris	Afrotheria	AHIN01144502.1	
	Tupaia belangeri	Boreoeutheria	AAPY01451866.1	
	Tursiops truncatus	Boreoeutheria	ABRN02551903.1	
				se nadaljuje

Con	Organizam	Talzanamalza	Štovilko zoporodio	Vromocomelro
Gen	Organizem	Taksonomska	Stevnka zaporeuja	Kromosomska
		skupina		юкастја
	Vicugna pacos	Boreoeutheria	ABRR01273839.1	
PNMA6				
	Ailuropoda melanoleuca	Boreoeutheria	XP_002930132.1	
	Canis lupus familiaris	Boreoeutheria	XP_549350	
	Canis lupus familiaris	Boreoeutheria	AAEX03027066.1	
	Dasypus novemcinctus	Xenarthra	AAGV03164002.1	
	Dipodomys ordii	Boreoeutheria	ABRO01263040	
	Dipodomys ordii	Boreoeutheria	ABRO01263040.1	
	Echinops telfairi	Afrotheria	AAIY01541208.1	
	Equus caballus	Boreoeutheria	ENSECAT00000014076	
	Felis catus	Boreceutheria	ACBE01511238	
	Homo sapiens	Boreceutheria	NP 116271	X (Xa28)
	Ictidomys tridecemlineatus	Boreceutheria	AGTP01098201 1	(1)
	Mustala putorius furo	Dorecoutheria	AEVD01114520.1	
	Musicia paiorias jaro	Doreceutieria	ADE02015504 1	
	Myous incijugus	Doreoeutieria	AAPE02013304.1	
		Boreoeutneria	AAQR03194462.1	
	Pongo abelii	Boreoeutheria	ABGA01401781.1	
	Pteropus vampyrus	Boreoeutheria	ABRP01291682	
	Trichechus manatus latirostris	Afrotheria	AHIN01144511.1	
	Tursiops truncatus	Boreoeutheria	ENSTTRP0000001277	
ZCCHC12				
	Ailuropoda melanoleuca	Boreoeutheria	XP_002916128.1	
	Bos taurus	Boreoeutheria	NP_001069665	
	Canis lupus familiaris	Boreoeutheria	XP_549212	
	Cavia porcellus	Boreoeutheria	XP_003467642.1	
	Cricetulus griseus	Boreoeutheria	AFTD01039914.1	
	Dasypus novemcinctus	Xenarthra	AAGV03098506.1	
	Dipodomys ordii	Boreoeutheria	ABRO01263293.1	
	Equus caballus	Boreoeutheria	XP_001487947	
	Erinaceus europaeus	Boreoeutheria	AANN01151507	
	Felis catus	Boreoeutheria	ACBE01504543.1	
	Gorilla gorilla	Boreoeutheria	CABD02404599.1	
	Homo sapiens	Boreoeutheria	NP_776159	X (Xq24)
	Ictidomys tridecemlineatus	Boreoeutheria	AAQQ01371946.1	
	Macaca mulatta	Boreoeutheria	XP_001099794	
	Microcebus murinus	Boreoeutheria	ABDC01181532	
	Mus musculus	Boreoeutheria	NP_082601	
	Mustela putorius furo	Boreoeutheria	AGTQ01053520.1	
	Nomascus leucogenys	Boreoeutheria	XP_003262290.1	
	Oryctolagus cuniculus	Boreoeutheria	XP_002/20256.1	
	Ovis aries	Boreoeutheria	ACIV012295118.1	
	Pan troglodytes	Boreoeutheria	XP_529128.3	
	Pongo abelii	Boreoeutheria	XP_002832079	
	Pongo abelii	Boreoeutheria	XP_002831991.1	
	Procavia capensis	Airotheria	ABRQ0121/588	
	Pteropus vampyrus	Boreoeutneria	ABRP01045922.1	
	Kallus horvegicus	Boreoeutneria	NP_001014087	
	Spermopnius triaecemineatus	Boreceutheric	AAQQUI3/1940 VD 002125207 1	
	Sus scioja Trichachus manatus latinostria	Afrotheria	AF_003133397.1 AHIN01072378 1	
	Tursions truppatus	Boreceutheria	ABRN02045262 1	
	Vicuona nacos	Boreceutheria	ABRR01371617 1	
PNMAL1	, cugna pacos	Dorocculiona	· DIMO13/101/.1	
A 1 1174/11/1	Ailuropoda melanoleuca	Boreceutheria	ACTA01042111 1	
	Ros taurus	Borecoutheric	ND 001005092	
	Callithrix jacobus	Boreceutheria	ACEV01175909 1	
	Carrier in jaccinas	201000unionu		

nadaljevanje preglednice P1. Seznam genov uporabljenih pri filogenomskih analizah

se nadaljuje

Gen	Organizem	Taksonomska skupina	Stevilka zaporedja	Kromosomska lokacija
	Canis lupus familiaris	Boreoeutheria	AAEX03000861.1	*
	Cavia porcellus	Boreoeutheria	XP_003464736.1	
	Dasypus novemcinctus	Xenarthra	AAGV03076085.1	
	Dipodomys ordii	Boreoeutheria	ABRO01140072.1	
	Equus caballus	Boreoeutheria	AAWR02030549.1	
	Gorilla gorilla	Boreoeutheria	CABD02181514.1	
	Heterocephalus glaber	Boreoeutheria	AFSB01041974.1	
	Homo sapiens	Boreoeutheria	NP_060685	19 (19q13.32)
	Ictidomys tridecemlineatus	Boreoeutheria	AGTP01089491.1	
	Loxodonta africana	Afrotheria	AAGU03011748.1	
	Macaca mulatt	Boreoeutheria	AEHK01141396.1	
	Microcebus murinus	Boreoeutheria	ABDC01329145	
	Mus musculus	Boreoeutheria	NP_001007570	
	Mustela putorius furo	Boreoeutheria	AEYP01068380.1	
	Myotis lucifugus	Boreoeutheria	AAPE02061731	
	Nomascus leucogenys	Boreoeutheria	ADFV01102146.1	
	Oryctolagus cuniculus	Boreoeutheria	AAGW02072742.1	
	Otolemur garnettii	Boreoeutheria	AAQR03153780.1	
	Pan troglodytes	Boreoeutheria	XP_524313	
	Pongo abelii	Boreoeutheria	NP_001125341	
	Pteropus vampyrus	Boreoeutheria	ABRP01162274.1	
	Rattus norvegicus	Boreoeutheria	NP 001101943	
	Tarsius syrichta	Boreoeutheria	ABRT010554846.1	
	Trichechus manatus latirostris	Afrotheria	AHIN01084034.1	
	Tursiops truncatus	Boreoeutheria	ABRN02000473.1	
PNMAL2	1			
	Callithrix jacchus	Boreceutheria	ACEV01108478 1	
	Echinons telfairi	Afrotheria	AAIY01565835 1	
	Homo saniens	Boreceutheria	NP 065760	19(19a1332)
	Lovodonta africana	Afrotheria	XP 003/067/5 1	1) (1)(1)(1).52)
	Macaca mulatta	Boreceutheria	A ANU 01182480 1	
	Mucaca matatta Mucamusculus	Doreceutheria	ND 001002106	
	Mustala putorius furo	Boreceutheria	AEVD01068377 1	
	Nomascus leucogenys	Boreceutheria	ADEV01102148 1	
	Atolemur garnettii	Boreceutheria	AAOR03153784	
	Dan troaledutes	Borecoutheria	VD 001167065	
	Pan irogioayies	Doreoeutieria	XF_001107905	
	Pongo abelli Butture u como cicco	Boreoeutheria	AP_002829490	
	Rattus norvegicus	Boreoeutheria	NP_001100951	
2001010	Trichechus manatus latirostris	Afrotheria	AHIN01084034.1	
ZCCHC18				
	Cricetulus griseus	Boreoeutheria	AFTD01070222.1	
	Homo sapiens	Boreoeutheria	NP_001137450	X (Xq22.2)
	Ictidomys tridecemlineatus	Boreoeutheria	AGTP01013369.1	
	Mus musculus	Boreoeutheria	NP 001030586	
	Nomascus leucogenvs	Boreceutheria	ADFV01007938.1	
	Canis lupus familiaris	Boreceutheria	XP 549160 3	
	Pan troglodytes	Boreceutheria	XP_003317651_1	
	Rattus norvagious	Borecoutheria	A AUX01112221 1	
	Trichachus manatus latinostiis	A froth and	AAHAUTTI3221.1	
CCDC	1 richechus manalus lattrostris	Airoineria	AHINU113408/.1	
CDC8				
	Bos taurus	Boreoeutheria	AAX31380.1	
	Canis lupus familiaris	Boreoeutheria	XP_003432615.1	
	Cavia porcellus	Boreoeutheria	XP_003464735.1	
	Equus caballus	Boreoeutheria	XP_001502968.3	
	Homo sapiens	Boreoeutheria	NP_114429.2	19 (19q13.32)
	Loxodonta africana	Afrotheria	AAGU03011749.1	
				se nadaljuje

Car	Organizari		Štarilla zazandia	Vacana
Gen	Organizem	Taksonomska	Steviika zaporedja	кготозотяка
		skupina		lokacija
	Macaca mulatta	Boreoeutheria	XP_001111790.1	
	Mus musculus	Boreoeutheria	EDL42061.1	
	Pan troglodytes	Boreoeutheria	XP 001167818.1	
	Pongo abelii	Boreceutheria	XP_002829489_1	
M DNIMA	I ongo ubeni	Boreoeduleriu	<u> 11 _002029409.1</u>	
	Macropus eugenii	Marsupialia	BAK55632.1	
	Monodelphis domestica	Marsupialia	AAFR03007941	3
GIN1				
	Ailuropoda melanoleuca	Boreoeutheria	ACTA01147224.1	
	Anolis carolinensis	Sauropsida	ENSACAP0000006979	
	Bos taurus	Boreceutheria	DAA22293 1	
	Callithrin jacchus	Boreceutheria	ACEV01006841 1	
	Cania lunua familiania	Dorecocutheria	ND 002620006 1	
	Canis iupus jamiliaris	Boreoeutieria	AP_003039000.1	
	Choloepus noffmanni	Xenarthra	ABVD01699732.1	
	Chrysemys picta bellii	Sauropsida	AHGY01271991.1	
	Dasypus novemcinctus	Xenarthra	ENSDNOP0000015601	
	Daubentonia madagascariensis	Boreoeutheria	AGTM011584885.1	
	Dipodomys ordii	Boreoeutheria	ENSDORP0000010277	
	Echinops telfairi	Afrotheria	AAIY01023878.1	
	Eauus caballus	Boreoeutheria	ENSECAT0000022350	
	Folis catus	Boreceutheria	ACBE01035028 1	
	Callus gallus	Sauropsida	ENSGAL T0000024621	7
	Ganilla serilla	Dauropsida Dauro a secto a si a	CADD02200754 1	L
	Gorilla gorilla	Boreoeutneria	CABD02299754.1	
	Heterocephalus glaber	Boreoeutheria	AFSB01096689.1	
	Homo sapiens	Boreoeutheria	NP_060146.2	5 (5q21.1)
	Ictidomys tridecemlineatus	Boreoeutheria	AGTP01043191.1	
	Loxodonta africana	Afrotheria	XP_003405049.1	
	Macaca mulatta	Boreoeutheria	NP_001181013	
	Macropus eugenii	Marsupialia	ABQO020290995.1	
	Meleagris gallopavo	Sauropsida	ADDD01043941.1	
	Melopsittacus undulatus	Sauropsida	AGAI01041143.1	
	hudoerioar	Suaroporau		
	Microcobus murinus	Boreceutheria	ABDC01264238 1	
	Monodolphis domostica	Margunialia	ENSMODT0000024502	
	Monodelphis domestica	Ransuplana Domo outhorio	ND 080526	
	WIUS MUSCUIUS	Boreoeutieria	NP_080320	
	Mustela putorius furo	Boreoeutheria	AEYP010/2838.1	
	Myotis lucifugus	Boreoeutheria	AAPE02000394.1	
	Nomascus leucogenys	Boreoeutheria	ADFV01187285.1	
	Ochotona princeps	Boreoeutheria	ENSOPRP00000013878	
	Ornithorhynchus anatinus	Monotremata	ENSOANP0000008669	
	Oryctolagus cuniculus	Boreoeutheria	XP_002713990.1	
	Otolemur garnettii	Boreoeutheria	AAQR03059801.1	
	Pan troglodytes	Boreoeutheria	XP 526970.2	
	Procavia capensis	Afrotheria	ABR001101301 1	
	Pteropus vampyrus	Boreceutheria	ENSPVAP0000005752	
	Pattus normagicus	Borecoutheria	ND 001020183 1	
		Managerialia	AEEX01220105.1	
	Sarcophilus narrisii	Marsupiana	AFE 101228897.1	7
	Taeniopygia guttata	Sauropsida	EINSTGUP0000000950	L
	Trichechus manatus latirostris	Atrotheria	AHIN01042750.1	
	Tursiops truncatus	Boreoeutheria	ENSTTRP0000000995	
GIN2				
	Callorhinchus milii	Chondrichthyes	AAVX01272723.1	
	Danio rerio	Actinopterygii	CABZ01030983.1	
	Gallus gallus	Sauropsida	ENSGALT00000005860	
	Latimeria chalumnae	Coelacanthimorpha	ENSLACT0000020312	
	Macronus eugenii	Marsunialia	ABOO020016960 1	
	Meleaaris gallonguo	Sauronsida	FNSMG&T0000003041	
	Sarcophilus harrisii	Marsunialia	ΔFFV01377517 1	
	Surcophilus nurristi	maisupiana	1 I I 1 01377317.1	sa nadalinia
				se nauaijuje

Com		Tal-aan amal-a	Štorillo zanonodio	Vuomosomalao
Gen	Organizem	Такзопотяка	Steviika zaporedja	Кготозотска
		skupina		lokacija
	Taeniopygia guttata	Sauropsida	ABQF01047089.1	
	Takifugu rubripes	Actinopterygii	CAAB02000649.1	
	Xenopus tropicalis	Amphibia	AAMC01042891.1	
GIN-like		•		
	Branchiostoma floridae	Cephalochordata	ABEP01009594	
	Branchiostoma floridae	Cephalochordata	ABEP0200003 1	
	Ciona intestinalia	Urochordoto	ENSCINT0000024800	
		Curlastamata	ENSCIN 100000034809	
	Petromyzon marinus	Cyclostomata	ENSPMA10000009777	
	Petromyzon marinus	Cyclostomata	ENSPMA10000009777	
SCAND3				
	Ailuropoda melanoleuca	Boreoeutheria	ACTA01148370.1	
	Bos taurus	Boreoeutheria	XP_002702938.1	
	Callithrix jacchus	Boreoeutheria	ACFV01153170.1	
	Canis lupus familiaris	Boreoeutheria	XP_545451.2	
	Choloepus hoffmanni	Xenarthra	ABVD01519353.1	
	Dasypus novemcinctus	Xenarthra	AAGV03013662.1	
	Dauhentonia madagascariensis	Boreceutheria	AGTM011946033 1	
	Equis caballus	Boreceutheria	XP 001504028 1	
	Equis cubulus Folio octus	Borecoutheria	ACDE01159425 1	
	Fells calus Conilla conilla	Doreoeutileria Doreoeutileria	ACBE01130433.1	
	Gorilla gorilla	Boreoeutheria	CABD02515941.1	c (c 22 1)
	Homo sapiens	Boreoeutheria	NP_443155.1	6 (6p22.1)
	Ictidomys tridecemlineatus	Boreoeutheria	AGTP01117005.1	
	Macaca mulatta	Boreoeutheria	XP_001101327.1	
	Macropus eugenii	Marsupialia	ABQO020574066.1	
	Mustela putorius furo	Boreoeutheria	AGTQ01007431.1	
	Nomascus leucogenys	Boreoeutheria	ADFV01069104.1	
	Otolemur garnettii	Boreoeutheria	AAQR03175422.1	
	Ovis aries	Boreoeutheria	ACIV011864315.1	
	Pan troglodytes	Boreceutheria	XP 527300.2	
	Pongo abelii	Boreceutheria	XP_002816663_1	
	Procavia canansis	Afrotheria	ABRO01339340 1	
	Snormonhilus tridocomlineatus	Porocouthorio	ADRQ01337340.1	
	Spermophilus triaecemineatus	Doreoeutheria	AAQQ01208141	
	Sus scroja	Boreoeutneria	AP_003128285.1	
	Trichechus manatus latirostris	Afrotheria	AHIN01123797.1	
	Tursiops truncatus	Boreoeutheria	ABRN02152968.1	
KRBA2				
	Ailuropoda melanoleuca	Boreoeutheria	ACTA01016012.1	
	Bos taurus	Boreoeutheria	XP_002702194.1	
	Callithrix jacchus	Boreoeutheria	XP 002806845.1	
	Canis lupus familiaris	Boreoeutheria	XP_003434716.1	
	Choloepus hoffmanni	Xenarthra	ABVD01164029.1	
	Dasynus novemcinctus	Xenarthra	AAGV03120505 1	
	Daubentonia madagascariensis	Boreceutheria	AGTM011803390 1	
	Equip achallus	Dorecocutheria	XD 001502262 2	
	Equus cubunus Eolia oatua	Doreceutheria	ACDE01571580 1	
		Doreoeutieria	ACBEU13/1369.1	
	Gorilla gorilla	Boreoeutheria	CABD02146056.1	
	Homo sapiens	Boreoeutheria	NP_998762.1	17 (17p13.1)
	Loxodonta africana	Afrotheria	AAGU03065723.1	
	Macaca mulatta	Boreoeutheria	XP_001113012.1	
	Microcebus murinus	Boreoeutheria	ABDC01211585.1	
	Mustela putorius furo	Boreoeutheria	AEYP01050843.1	
	Myotis lucifugus	Boreoeutheria	AAPE02045206.1	
	Nomascus leucogenvs	Boreoeutheria	XP_003274599.1	
	Ovis aries	Boreoeutheria	ACIV011623187.1	
	Pan troolodytes	Boreceutheria	AACZ03107953 1	
	Pongo abelii	Boreceutheria	NP 001127348 1	
	Saroonhilus harrisii	Marcupialia	AEEK01165105 1	
	Sus sorofa	Boreceutheric	VD 002258222 1	
	sus scroju	Doreoeumeria	AI _003330333.1	aa madal::-
				se nadaljuje

Com	Organizam	Talsaanamalsa	Štovilko zoporodio	Vromocomelro
Gen	Organizem	Такзопотяка	Stevnka zaporedja	Кготозотска
		skupina		lokacija
	Trichechus manatus latirostris	Afrotheria	AHIN01050211.1	
	Trichosurus vulpecula	Marsupialia	DY605310.1	
	Vicugna pacos	Boreoeutheria	ABRR01279825.1	
NYNRIN				
	Homo sapiens	Boreoeutheria	NP_079357.2	14 (14q12)
	Mus musculus	Boreoeutheria	NP 001035161.1	
	Pongo abelii	Boreoeutheria	XP_002824664.1	
	Pan troglodytes	Boreoeutheria	XP_509877.2	
	Macaca mulatta	Boreoeutheria	XP_001104789.2	
	Fauus caballus	Boreceutheria	ENSEC AP0000014240	
	Canis lupus familiaris	Boreoeutheria	XP 547747.3	
	Orvetolagus cuniculus	Boreceutheria	XP 002718154 1	
	Bos taurus	Boreceutheria	NP_001179388_1	
	Monodelphis domestica	Marsunialia	XP_001380281_1	
	Lorodonta africana	Afrotheria	XI_001380281.1 XI _003421016.1	
	Cavia poreellus	Porocouthorio	XI_003474221 1	
	Muotia luoifuqua	Boreceutheria	AF_0034/4331.1 ENSMLUD0000010277	
	Trick ochora man atua latino atuia	Afrethania	ATHNO1147228 1	
	Trichechus manatus tattrostris	Afrotheria	ADD001414840 1	
	Procavia capensis	Alrotheria	ABRQ01414840.1	
	Echinops telfairi	Afrotheria	AAIY01153//1.1	
	Dasypus novemcinctus	Xenarthra	AAGV03121121.1	
	Choloepus hoffmanni	Xenarthra	ABVD018/2/89.1	
	Sarcophilus harrisii	Marsupialia	AEFK01009009.1	
	Macropus eugenii	Marsupialia	ABQO020422001.1	
	Pteropus vampyrus	Boreoeutheria	ABRP01121463.1	
	Felis catus	Boreoeutheria	AANG02157131.1	
	Tursiops truncatus	Boreoeutheria	ABRN02279328.1	
	Mustela putorius furo	Boreoeutheria	AGTQ01027410.1	
	Ailuropoda melanoleuca	Boreoeutheria	ACTA01019458.1	
	Gorilla gorilla	Boreoeutheria	CABD02105976.1	
	Callithrix jacchus	Boreoeutheria	XP_002753733.1	
	Nomascus leucogenys	Boreoeutheria	XP_003260748.1	
	Ictidomys tridecemlineatus	Boreoeutheria	AGTP01099316.1	
	Otolemur garnettii	Boreoeutheria	AAQR03175878.1	
	Microcebus murinus	Boreoeutheria	ABDC01231164.1	
	Heterocephalus glaber	Boreoeutheria	EHB07434.1	
	Tupaia belangeri	Boreoeutheria	AAPY01734038.1	
ASPRV1	1 0			
1101 11 1	Ailuropoda melanoleuca	Boreceutheria	ACTA01176775 1	
	Ros taurus	Boreceutheria	DAAA02031529 1	
	Callithrix jacchus	Boreceutheria	ACEV01110198 1	
	Canis lunus familiaris	Boreceutheria	A A E X 03007628 1	
	Cavia porcellus	Boreceutheria	A A K NO2036909 1	
	Cholognus hoffmanni	Vanarthra	ARVD01643553	
	Choicepus nojjmanni	Vanarthra	ABVD01045555	
	E avera e al allera		AAU V05145024.1	
	Equus cabanus Estis estas	Doreoeutheria	AAWK02005757.1	
	<i>Certilla e crilla</i>	Doreoeutieria	AANG02145111 CADD02102078 1	
	Gorilla gorilla	Boreoeutneria	CABD02193978.1	
	Heterocephalus glaber	Boreoeutneria	AFSB01223045.1	0 (0, 10, 0)
	Homo sapiens	Boreoeutneria	NP_690005.2	2 (2p13.3)
	Ictidomys tridecemlineatus	Boreoeutheria	AAQQ01029984.1	
	Loxodonta africana	Afrotheria	AAGU03051431.1	
	Macaca mulatta	Boreoeutheria	AANU01291061.1	
	Macropus eugenii	Marsupialia	ABQ0021096366.1	
	Microcebus murinus	Boreoeutheria	ABDC01048367.1	
	Monodelphis domestica	Marsupialia	AAFR03004014.1	
	Mus musculus	Boreoeutheria	NP_080690.2	
	Mustela putorius furo	Boreoeutheria	AEYP01010392.1	
				se nadaljuje

Gen	Organizem	Taksonomska	Številka zaporedja	Kromosomska
		skupina		lokacija
	Myotis lucifugus	Boreoeutheria	AAPE02025611.1	
	Nomascus leucogenys	Boreoeutheria	ADFV01010511.1	
	Ochotona princeps	Boreoeutheria	AAYZ01083901	
	Ornithorhynchus anatinus	Monotremata	AAPN01199678.1	
	Oryctolagus cuniculus	Boreoeutheria	AAGW02006848.1	
	Otolemur garnettii	Boreoeutheria	AAQR03102198.1	
	Pan troglodytes	Boreoeutheria	AADA01178871.1	
	Pongo abelii	Boreoeutheria	ABGA01353826.1	
	Pteropus vampyrus	Boreoeutheria	ABRP01166574.1	
	Rattus norvegicus	Boreoeutheria	EDL91253.1	
	Sarcophilus harrisii	Marsupialia	AFEY01388655.1	
	Tursiops truncatus	Boreoeutheria	ABRN02331053.1	
ARC				
	Ailuropoda melanoleuca	Boreoeutheria	ACTA01148197.1	
	Anolis carolinensis	Sauropsida	AAWZ02020354.1	
	Bos taurus	Boreoeutheria	NP_001193336.1	
	Callithrix jacchus	Boreoeutheria	ACFV01185911.1	
	Canis lupus familiaris	Boreoeutheria	AACN010440983.1	
	Cavia porcellus	Boreoeutheria	AAKN02023196.1	
	Chrysemys picta bellii	Sauropsida	AHGY01086810.1	
	Dasypus novemcinctus	Xenarthra	AAGV03311726.1	
	Gallus gallus	Sauropsida	NP 989763.1	
	Gorilla gorilla	Boreoeutheria	CABD02372626.1	
	Heterocephalus glaber	Boreoeutheria	AFSB01232674.1	
	Homo sapiens	Boreoeutheria	NP_056008.1	8 (8q24.3)
	Ictidomys tridecemlineatus	Boreoeutheria	AGTP01001239.1	
	Loxodonta africana	Afrotheria	AAGU03089166.1	
	Macaca mulatta	Boreoeutheria	AANU01107068.1	
	Macropus eugenii	Marsupialia	ABOO020225819.1	
	Meleagris gallopavo	Sauropsida	XP 003205366.1	
	Melopsittacus undulatus	Sauropsida	AGAI01032183.1	
	budgerigar	1		
	Monodelphis domestica	Marsupialia	AAFR03005662 1	
	Mus musculus	Boreceutheria	NP 061260 1	
	Mustela putorius furo	Boreoeutheria	AEYP01014255.1	
	Nomascus leucogenys	Boreoeutheria	ADFV01097265.1	
	Ornithorhynchus anatinus	Monotremata	ΔΔΡΝΟ1117397 1	
	Otolemur garnettii	Boreoeutheria	AAOR03184588 1	
	Pan tragladytes	Boreceutheria	XP 528245 2	
	Pongo abelii	Boreceutheria	ABGA01006600 1	
	Procavia capensis	Afrotheria	ABR001626709 1	
	Pteropus vampyrus	Boreceutheria	ABRP01297506 1	
	Python molurus	Sauropsida	AEQU010333076.1	
	Rattus norvegicus	Boreoeutheria	NP 062234.1	
	Saimiri boliviensis	Boreoeutheria	AGCE01132459.1	
	Sarcophilus harrisii	Marsunialia	AFEY01455681	
	Sorex araneus	Boreoeutheria	AALT01329681 1	
	Taenionvoja outtata	Sauronsida	ABOF01048364 1	
	Trichechus manatus latirostris	Afrotheria	AHIN01125499 1	
	Tursions truncatus	Boreoeutheria	ABRN02468085.1	
	Xenopus tropicalis	Amphibia	AAMC01034049.1	

## PRILOGA B

# Shematski prikaz ohranjenosti sintenije udomačenih genov: (a) družina PNMA, (b) integraznih genov, (c) gena ARC in (d) gena ASPRV1.

**(a)** 





(c)





## PRILOGA C

# Vezavna mesta transkripcijskih faktorjev v promotorjih udomačenih genov pri človeku in miši.

### Udomačeni geni družine Sushi:

#### LDOC1:

Človek:: AP-1, ATF-2, AP-2gamma, FOXL1, Egr-4, COMP1, AP-2beta, c-Jun, AP-2alpha, AP-2alphaA

Miš: Elk-1, Meis-1, Meis-1b, HOXA9, HOXA9B, Nkx5-1, SEF-1(1), p300, POU3F2, Nkx6-1

#### LDOC1L:

Človek:: NF-1, Tal-1, CUTL1, Tal-1beta, E47, NRF-2, COMP1, STAT3, ITF-2

Miš: RFX1, PPAR-gamma1, PPAR-gamma2, HOXA9, HOXA9B, Meis-1, Meis-1a, c-Myb, Meis-1b, POU3F2

#### RGAG4:

Človek:: E2F-4, E2F-3a, E2F-5, E2F-2, Egr-4, MZF-1, E2F, E2F-1, COMP1, Pax-4a

Miš: XBP-1, Cdc5, IRF-1, PPAR-gamma1, PPAR-gamma2, HTF, CUTL1, NCX/Ncx, HEN1, GR

#### FAM127A:

Človek:: Elk-1, Nkx5-1, c-Ets-1

Miš: -

#### FAM127C:

Človek:: Elk-1, Nkx5-1, c-Ets-1

Miš: -

#### **PEG10:**

Človek:: C/EBPbeta

Miš: -

#### ZCCHC5:

Človek:: Sox5, PPAR-gamma1, Olf-1, FOXO1a, PPAR-gamma2, HNF-3beta, Arnt, FOXO1

Miš: POU3F2, POU3F2 (N-Oct-5a), POU3F2 (N-Oct-5b), POU2F1, POU2F1a, Bach1, Pax-6, Brachyury, Evi-1, Roaz

#### ZCCHC16:

Človek:: Sp1, RREB-1, NF-YA, HNF-3beta, NF-YB, CBF-A, CBF-B, CP1A, NF-Y, CBF(2)

Miš: PPAR-gamma1, PPAR-gamma2, C/EBPalpha, LCR-F1, RREB-1, HNF-4alpha1, HNF-4alpha2, FOXD1, NF-E2, NF-E2 p45

#### RTL1:

Človek:: AML1a, Pax-5, Olf-1, MyoD, E4BP4, C/EBPalpha, FOXJ2 (long isoform), FOXJ2

Miš: FOXC1, ARP-1, HFH-1, RP58, STAT1, STAT1alpha, STAT1beta, STAT2, STAT3, STAT4

#### RGAG1:

Človek:: STAT5B, AP-4, CUTL1, Evi-1, MEF-2A, SRY, POU2F1, POU2F1a, aMEF-2, RSRFC4

Miš: LCR-F1, GCNF-2, GCNF, GCNF-1, POU3F2, aMEF-2, MEF-2A, MZF-1, PPARalpha, Roaz

#### *C22orf29*:

Človek:: E2F-4, E2F-3a, E2F-5, E2F-2, LCR-F1, C/EBPalpha, CHOP-10, E2F, E2F-1, Pax-4a

Miš: -

#### Udomačeni geni družine PNMA

#### PNMA1:

Človek:: E2F-3a, E2F-4, E2F-5, Brachyury, HSF1 (long), E2F-2, E2F-1, E2F, HSF1short, ATF

Miš: ZIC2/Zic2, Roaz, Nkx3-1v1, Nkx3-1v2, Nkx3-1v3, Nkx3-1v4, Nkx3-1, Msx-1, Zic1, RFX1

#### PNMA2:

Človek:: AML1a, Pax-5, MyoD, Lmo2, AP-4, GATA-1, Egr-4, FOXL1, HEN1

Miš: NRSF form1, NRSF form2, GATA-1, ITF-2, Tal-1beta, HSF1 (long), HSF1 (short), YY1, MyoD, C/EBPalpha

#### PNMA3:

Človek:: E2F-3a, E2F-4, E2F-5, SREBP-1c, E2F-2, SREBP-1b, E2F-1, E2F, SREBP-1a, HOXA5

Miš: ATF2, CRE-BP1, ATF, Ik-3, c-Jun, RP58, NF-kappaB1, Roaz, ISGF-3, p53

#### MOAP1:

Človek:: AML1a, ATF-2, NF-kappaB, FOXL1, AREB6, IRF-2, NF-kappaB2, Meis-1a, NF-kappaB1, RSRFC4

Miš: Evi-1, ZID, c-Myb, STAT3, POU3F1, FOXJ2, FOXJ2 (long isoform), Nkx3-1, Nkx3-1 v1, Nkx3-1 v2
### PNMA5:

Človek:: COUP-TF1, LHX3b/Lhx3b, C/EBPbeta, AML1a, HNF-4alpha2, HNF-4alpha1, GATA-2, COUP-TF, CP2, LHX3a/Lhx3a

Miš: Nkx6-1, Evi-1, HSF1 short, HSF1 (long), Nkx2-5, NF-kappaB1, Pax-2, Pax-2a, SREBP-1a, SREBP-1b

### PNMA6A:

Človek:: CREB, SREBP-1a, deltaCREB, SREBP-1c, MyoD, NRSF form 2, SREBP-1b

Miš: -

### PNMAL1:

Človek:: ER-alpha, USF1, USF2, USF-1:USF-2, GATA-1, POU2F1, POU2F1b, POU2F1a, USF-1, POU2F1c

Miš: -

### PNMAL2:

Človek:: AhR, NF-1, Pax-5, Lmo2, Arnt, E47, AREB6, STAT3, Hand1, MRF-2

Miš: -

### ZCCHC12:

Človek:: GR, GR-beta, Nkx2-2, Meis-1b, POU2F1, POU2F1a, c-Myb, GR-alpha, Meis-1a, Meis-1

Miš: POU2F1, POU2F1a, POU2F2, POU2F2 (Oct-2.1), POU2F2B, POU2F2C, Oct-B1, oct-B2, oct-B3, Pax-3

#### ZCCHC18:

Človek:: STAT5B, AP-1, XBP-1, STAT5A, NF-E2 p45, POU2F1, GATA-6, POU2F1a, Zic3, NF-E2

Miš: Pax-3, GATA-1, Olf-1, GR, GR-alpha, CBF(2), CBF-A, CBF-B, CBF-C, CP1A

## CCDC8:

Človek:: NRSF form 1, Nkx2-5, NRSF form 2, AP-2gamma, Ik-2, HSF2, AP-2beta, AP-2alpha, Hlf, AP-2alphaA

Miš: -

#### ARC:

Človek:: Sp1, NRSF form 1, STAT5A, NRSF form 2, NF-kappaB, NF-kappaB1

Miš: SRF, SRF (504 AA), HOXA9B, Meis-1, Meis-1b, HOXA9, SREBP-1a, SREBP-1b, SREBP-1c, E47

## Proteazni udomačeni gen:

ASPRV1:

Človek:: Max1, AML1a, MAZR, NCX/Ncx, Pax-2, Pax-2a, LCR-F1, ATF6, c-Myc

Miš: ARP-1, FOXD1, HNF-1, HNF-1A, SEF-1 (1), ROR-alpha2, NF-kappaB, NF-kappaB1, RelA, Lmo2

# Udomačeni geni nastali iz integrazne domene:

# GIN1:

Človek:: AML1a, POU3F2 (N-Oct-5a), POU3F2 (N-Oct-5b), CUTL1, POU3F2, CREB, deltaCREB, Meis-1a, Meis-1

Miš: -

## NYNRIN (KIAA1305):

Človek:: ISGF-3, Ik-3, HEN1, Evi-1, GATA-1, GATA-2, GATA-3, LyF-1, HNF-4alpha1, CREB

Miš: -

# KRBA2:

Človek:: HOXA3, AML1a, GCNF, GATA-1, C/EBPalpha, AREB6, Nkx6-1, GCNF-1, COMP1, IRF-7A

Miš: -

# SCAND3:

Človek:: RFX1, NRSF form 1, Olf-1, NRSF form 2, TGIF, MRF-2, Roaz

Miš: -