

Genesis and Regulatory Wiring of Retroelement-Derived Domesticated Genes: A Phylogenomic Perspective

Janez Kokošar and Dušan Kordiš*

Department of Molecular and Biomedical Sciences, Josef Stefan Institute, Ljubljana, Slovenia

*Corresponding author: E-mail: dusan.kordis@ijs.si.

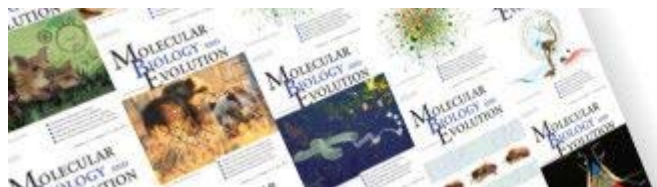
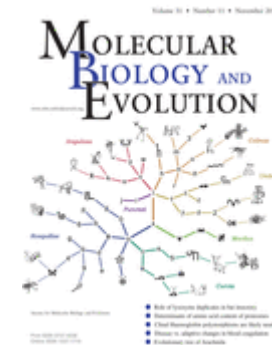
Associate editor: Jianzhi Zhang

Abstract

Molecular domestications of transposable elements have occurred repeatedly during the evolution of eukaryotes. Vertebrates, especially mammals, possess numerous single copy domesticated genes (DGs) that have originated from the intronless multicopy transposable elements. However, the origin and evolution of the retroelement-derived DGs (RDDGs) that originated from Metaviridae has been only partially elucidated, due to absence of genome data or to limited analysis of a single family of DGs. We traced the genesis and regulatory wiring of the Metaviridae-derived DGs through phylogenomic analysis, using whole-genome information from more than 90 chordate genomes. Phylogenomic analysis of these DGs in chordate genomes provided direct evidence that major diversification has occurred 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. Analysis of syntenic loci has shown that diverse RDDGs and their chromosomal positions were fully established in the ancestor of placental mammals. By analysis of active Metaviridae lineages in amniotes, we have demonstrated that RDDGs originated from retroelement remains. The chromosomal gene movements of RDDGs were highly dynamic only in the ancestor of placental mammals. During the domestication process, de novo acquisition of regulatory regions is shown to be a prerequisite for the survival of the DGs. The origin and evolution of de novo acquired promoters and untranslated regions in diverse mammalian RDDGs have been explained by comparative analysis of orthologous gene loci. The origin of placental mammal-specific innovations and adaptations, such as placenta and newly evolved brain functions, was most probably connected to the regulatory wiring of DGs and their rapid fixation in the ancestor of placental mammals.

Key words: molecular domestication, retroelement, phylogenomics, placentals, neofunctionalization, regulatory evolution.

Nastanek in regulatorna evolucija domesticiranih genov



MOLECULAR
BIOLOGY AND
EVOLUTION

IMPACT FACTOR
14.308!

MOLEKULARNA DOMESTIKACIJA

Proteini kodirani v transpozicijskih elementih (TE) omogočajo njihovo **izrezovanje, replikacijo in integracijo** v genomu. Tako bogat nabor lastnosti in aktivnosti je zanimiv in uporaben za gostiteljske celice, predvsem iz vidika evlucijskih inovacij. **Molekularna domestikacija TE predstavlja pomembno evlucijsko inovacijo, s katero so organizmi pridobili nove funkcionalne elemente.** Domesticirani geni so pogosto pridobili nove, od prvotne funkcije neodvisne vloge.

Ancestralni transpozicijski element (TE) Domesticiran gen

Številne kopije v genomu

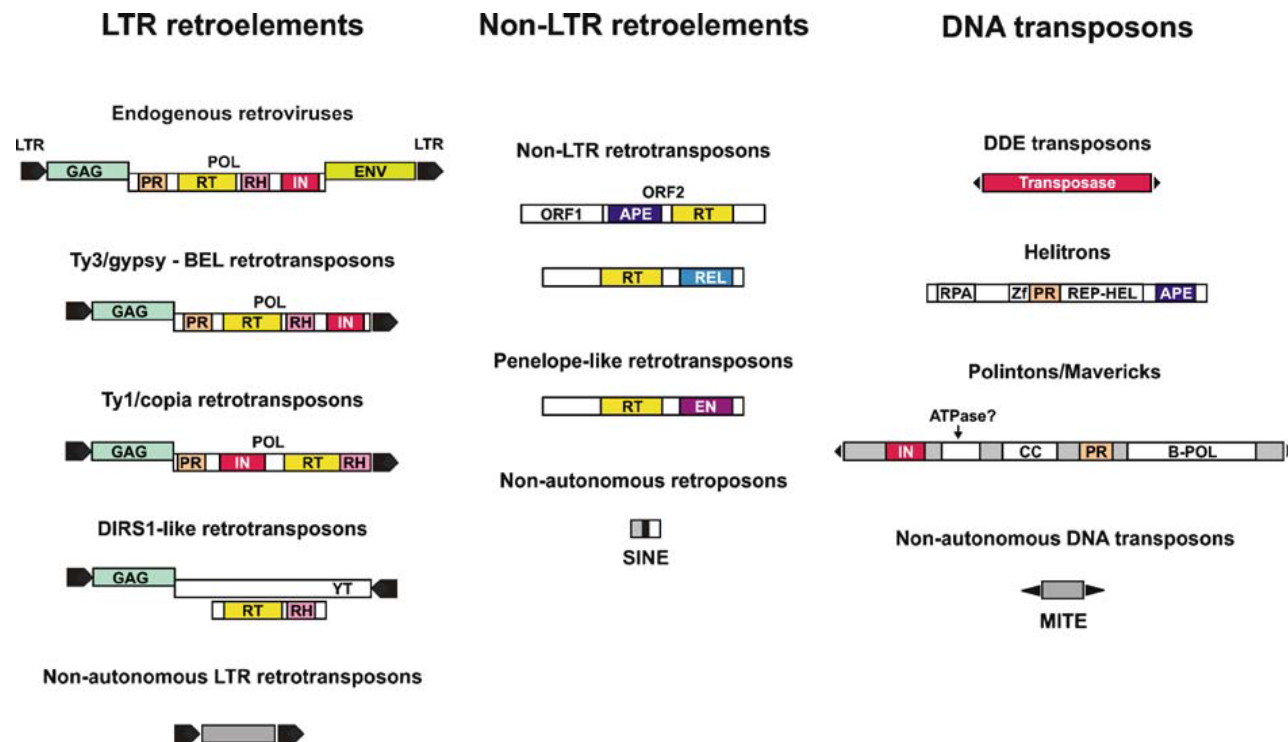
Ena sama kopija gena v genomu

Sposobnost avtonomne transpozicije

Ne morejo se premikati po genomu

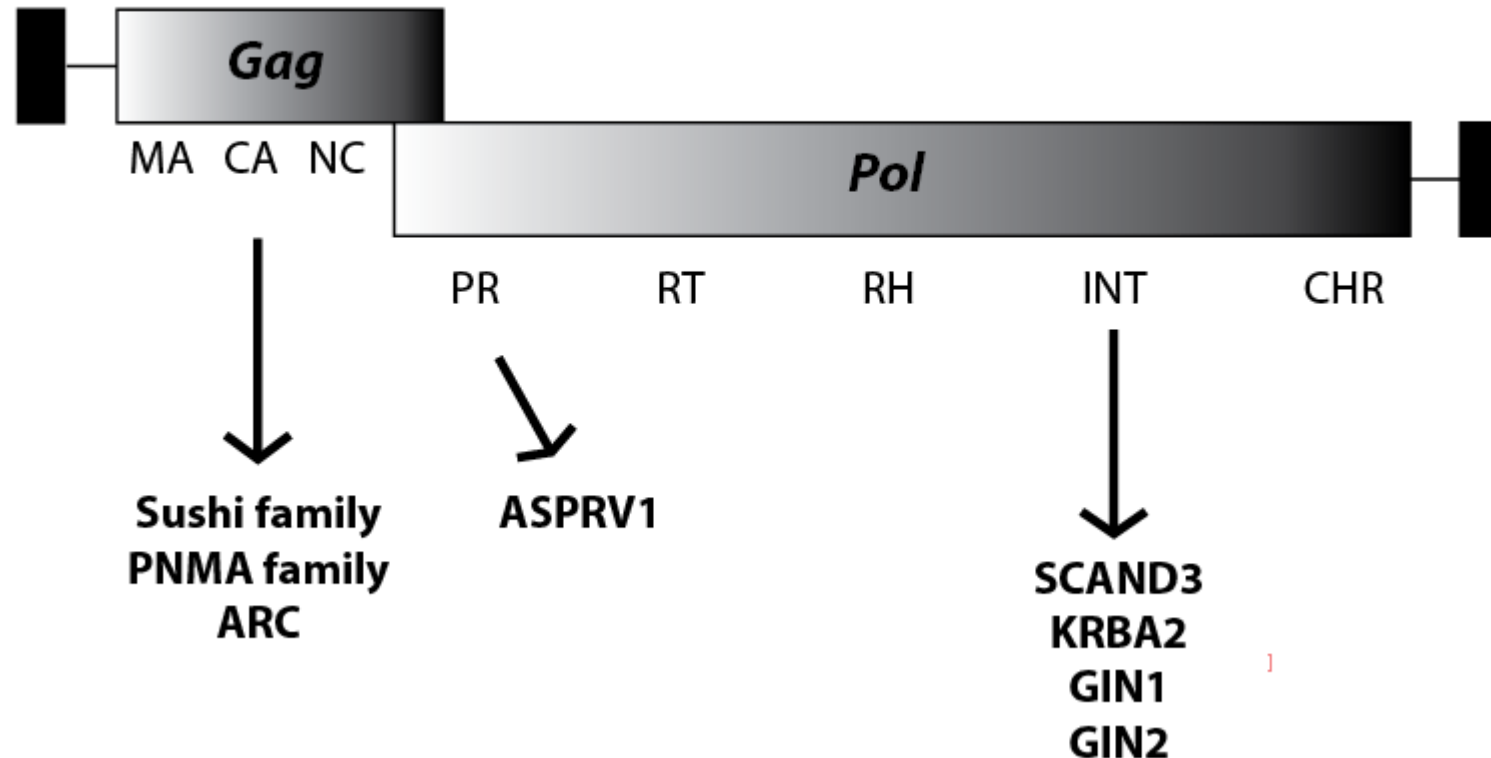
Prisotni na različnih mestih v genomu

Prisotni na ohranjenih ortolognih genskih lokusih



DRUŽINE DOMESTICIRANIH GENOV

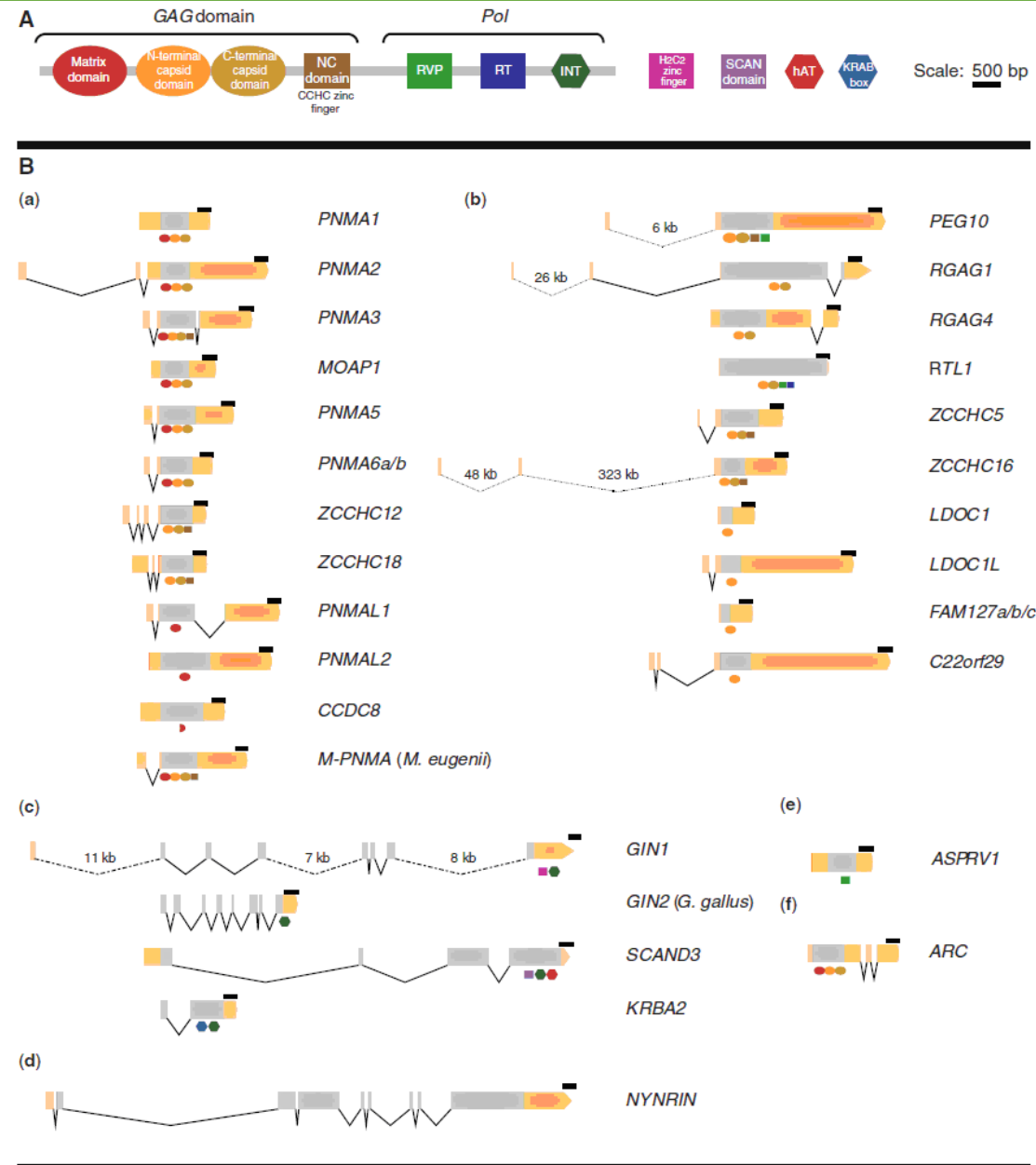
Analizirali smo **domesticirane gene (29 genov)** ki so nastali iz LTR retrotranspozonov [Metaviridae (Ty3/Gypsy)].



NAMEN RAZISKAV

- Pojasniti **nastanek, razširjenost, raznolikost in evolucijo domesticiranih genov pri vretenčarjih.**
- S pomočjo **filogenomske analize** pojasniti **genezo in močno povečanje števila DG** (analiza >90 vretenčarskih genomov).
- Pojasniti **časovni okvir domestikacij: kje in kdaj** so se metaviridne domene (gag, proteaza in integraza) transformirale v domesticirane gene.
- Pojasniti **mehanizme molekularne domestikacije ter *de novo* pridobivanja regulatornih zaporedij** teh genov, ki so pomembni za razumevanje njihove **funkcionalne diverzifikacije in pridobivanje novih bioloških vlog.**

VELIKE RAZLIKE V GENSKI STRUKTURI DOMESTICIRANIH GENOV



Gene structures of human RDDGs

(A) Color-coded **protein domains** of Metaviridae retroelements and additional domains that are present in human RDDG genes.

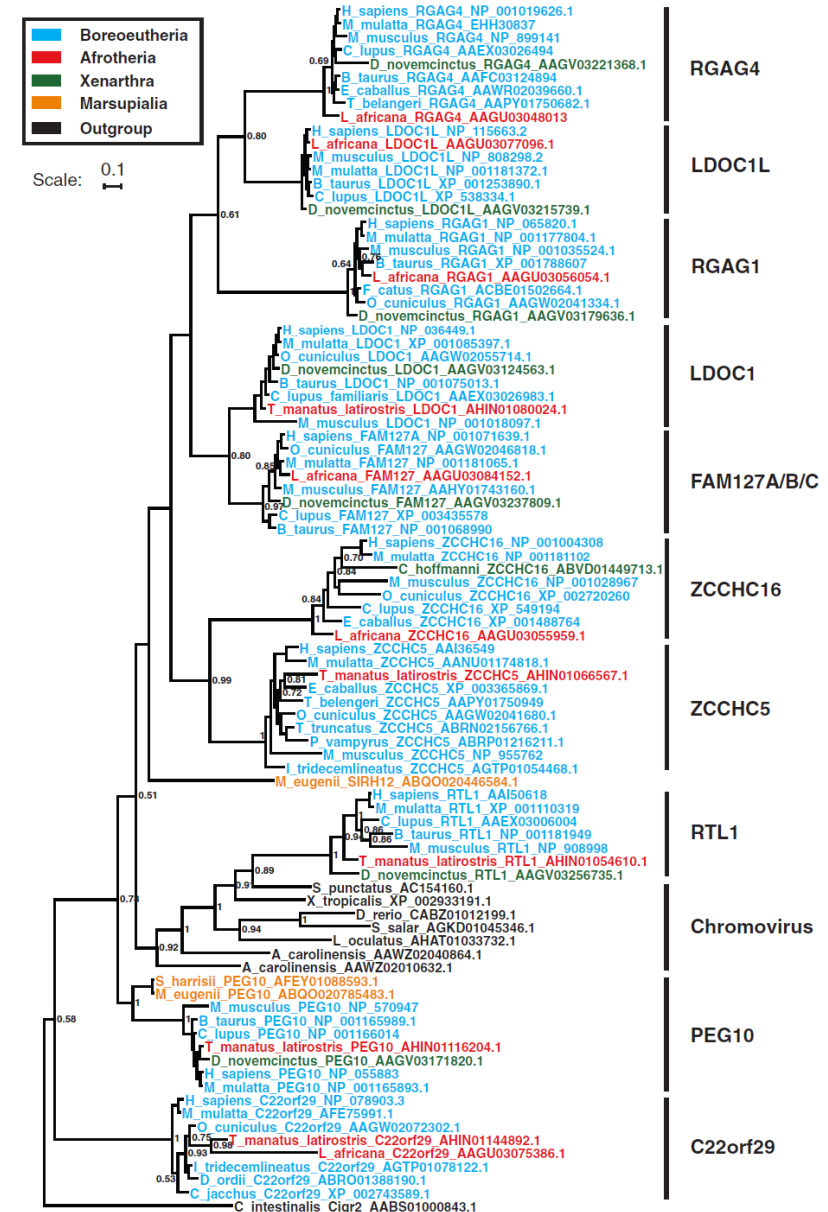
(B) **Exon–intron organization** of human RDDG genes. **Exons** are shown as boxes and **introns** as connecting lines. **Gray regions of exons** denote the protein coding sequences, whereas **UTRs** are represented as orange boxes.

Z analizo strukturne organizacije domesticiranih genov smo dokazali obsežen de novo „intron gain“ v predniku placentalnih sesalcev (Kordiš, Biology Direct 2011).

NEODVISNE MOLEKULARNE DOMESTIKACIJE

- Ločitev paralogov in ortologov
- Našli smo progenitorje DG
- Pojasnili smo evlucijske odnose znotraj genskih družin

- **FILOGENETSKA ANALIZA (družina SUSHI (Chromovirus))**
- MrBayes tree under Poisson+G4 model
- Alignment of the N-terminal capsid of gag domain.
- The scale bar corresponds to 0.1 substitutions per site.
- *C. intestinalis* Cigr2 retroelement was used to root this tree.

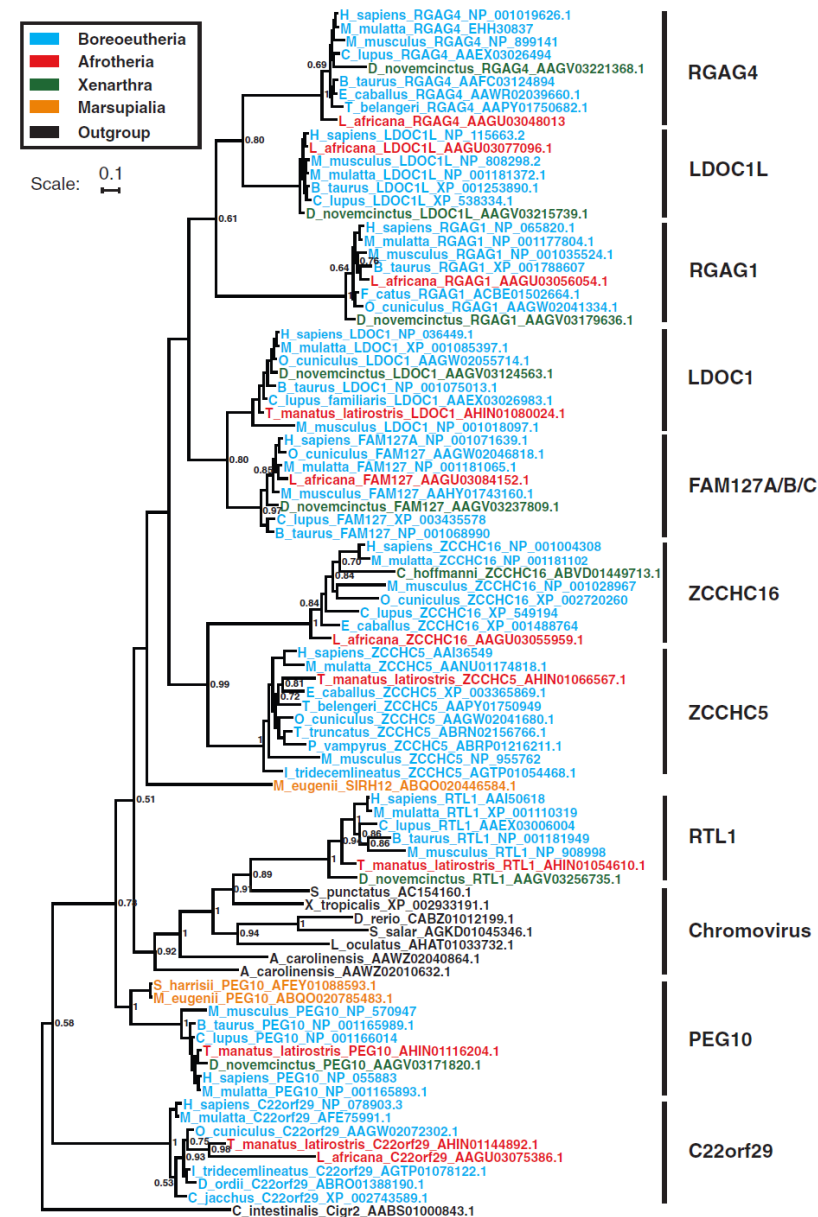


NEODVISNE MOLEKULARNE DOMESTIKACIJE

- Ločitev paralogov in ortologov
- Našli smo progenitorje DG
- Pojasnili smo evlucijske odnose znotraj genskih družin

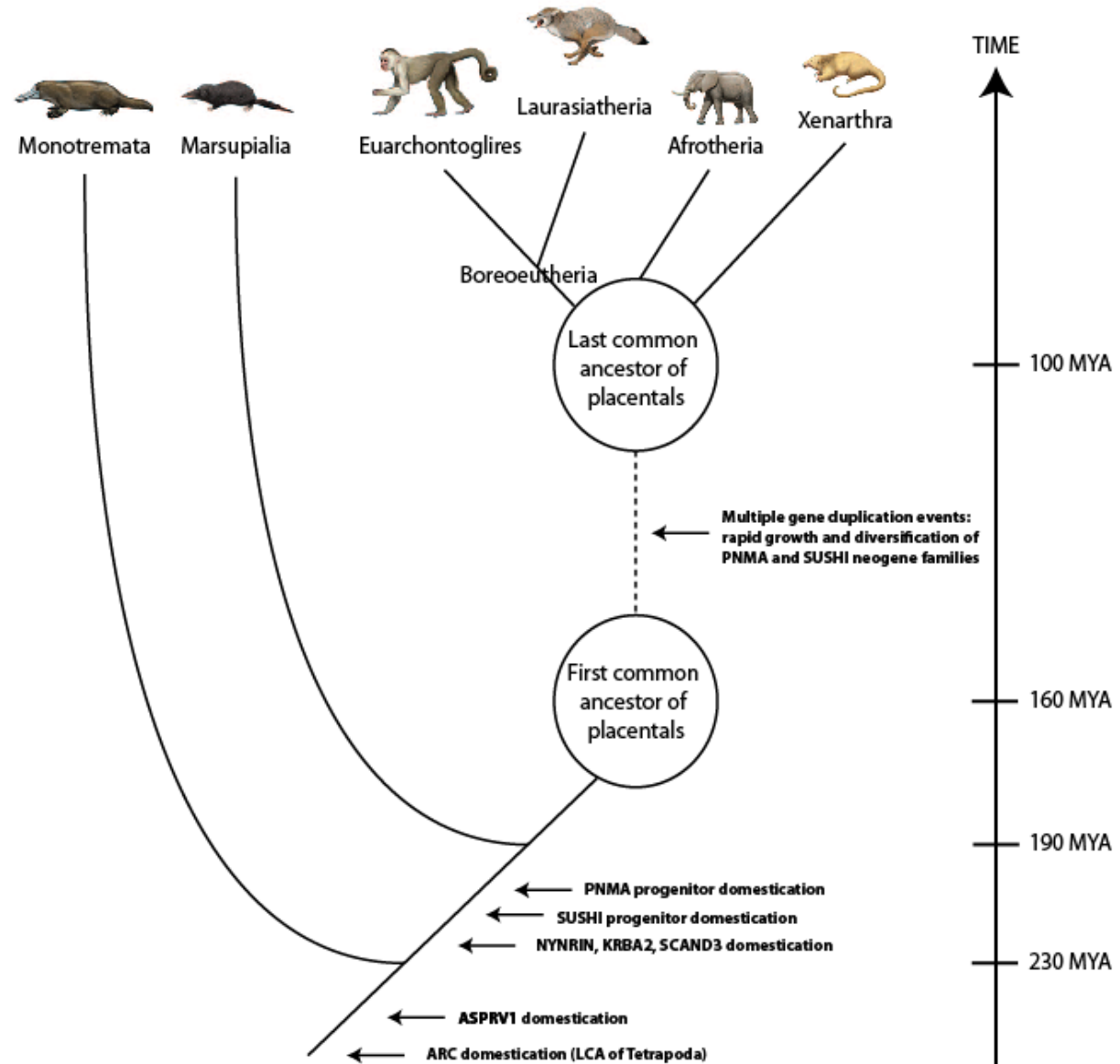
Domesticirani geni so nastajali preko več neodvisnih dogodkov molekularne domestikacije, nastale genske družine pa so se povečevale z genskimi duplikacijami.

- **FILOGENETSKA ANALIZA (družina SUSHI (Chromovirus))**
- MrBayes tree under Poisson+G4 model
- Alignment of the N-terminal capsid of gag domain.
- The scale bar corresponds to 0.1 substitutions per site.
- *C. intestinalis* Cigr2 retroelement was used to root this tree.

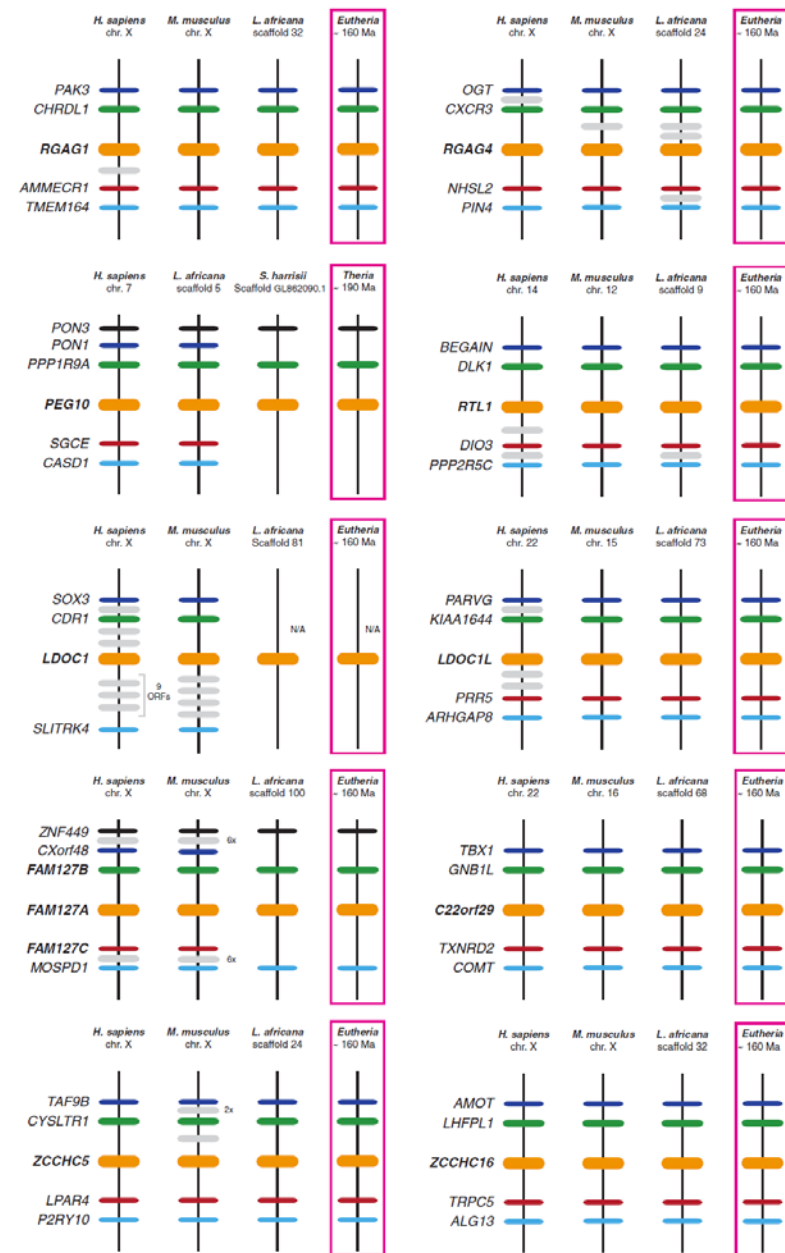


GENEZA DOMESTICIRANIH GENOV

Večina domesticiranih genov je nastala v predniku sesalcev

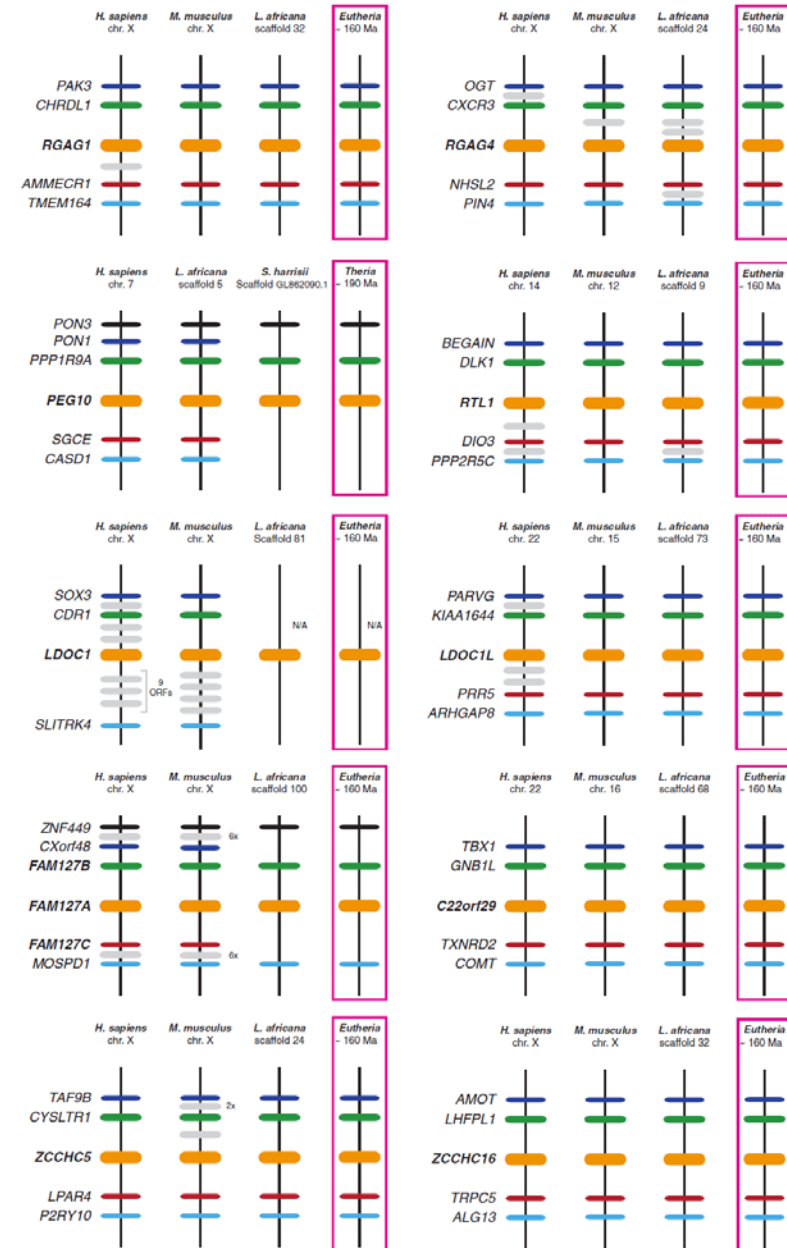


KROMOSOMSKI POLOŽAJI DOMESTICIRANIH GENOV



KROMOSOMSKI POLOŽAJI DOMESTICIRANIH GENOV

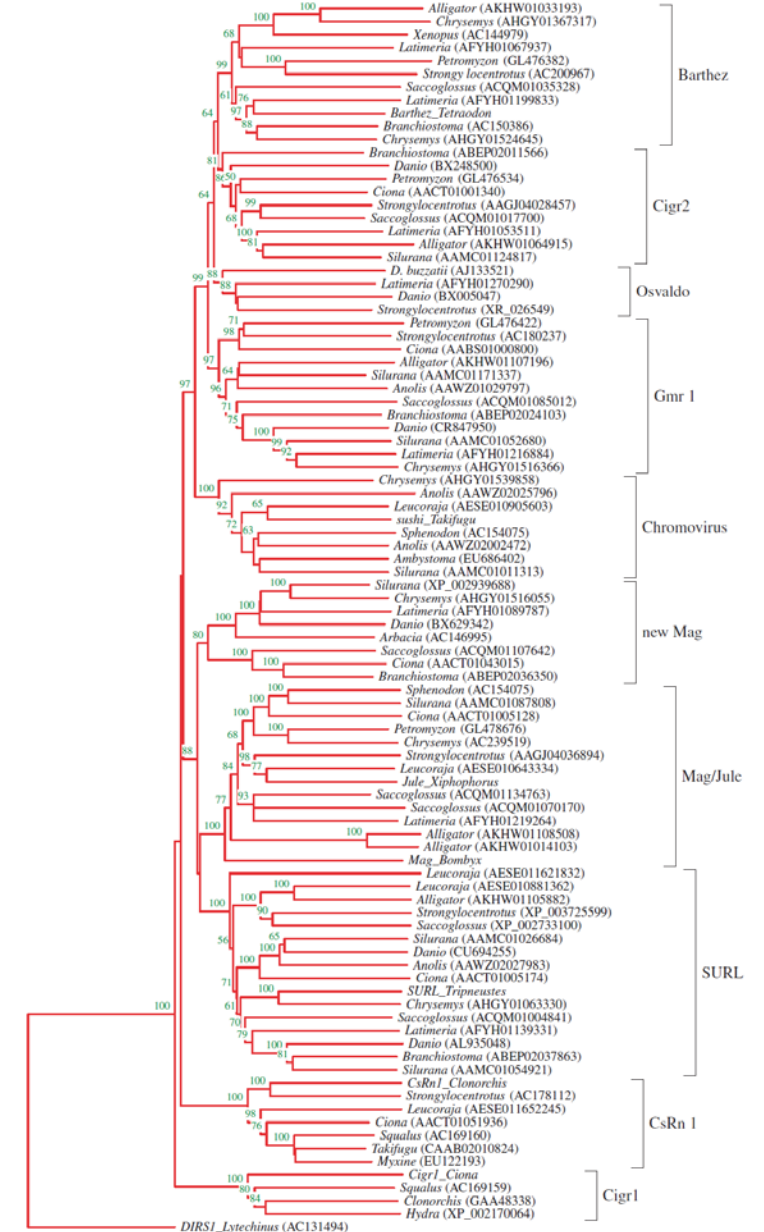
- S pomočjo **analize ohranjene sintenije** smo ugotovili „ancestral state“ **genomskih lokacij domesticiranih genov**.
- Analiza ohranjene sintenije je dokazala, da so bili **različni domesticirani geni in njihovi položaji na kromosomih prisotni že v predniku placentalnih sesalcev**.



DOMESTICIRANI GENI SO NASTALI IZ OSTANKOV RETROELEMENTOV

Metaviridae clade	Sauropsida				Mammals		Ancestral states			
	Lepidosauria	Turtles	Archosauria (crocodiles)	Archosauria (birds)	Prototheria	Theria	Sauropsida	Synapsida	Amniota	Tetrapoda
Chromovirus	+	*	*	Δ	Δ	Δ	+	+	+	+
Cigr2	Δ	Δ	+	Δ	Δ	Δ	+	+	+	+
Barthez	Δ	+	+	Δ	Δ	Δ	+	+	+	+
Gmr1	+	+	+	Δ	Δ	Δ	+	+	+	+
new Mag	Δ	+	Δ	Δ	Δ	Δ	+	+	+	+
SURL	+	+	+	Δ	Δ	Δ	+	+	+	+
Mag	+	+	+	Δ	Δ	Δ	+	+	+	+

+: full length elements present, Δ: elements absent, *: corrupted elements (retroelement remains)



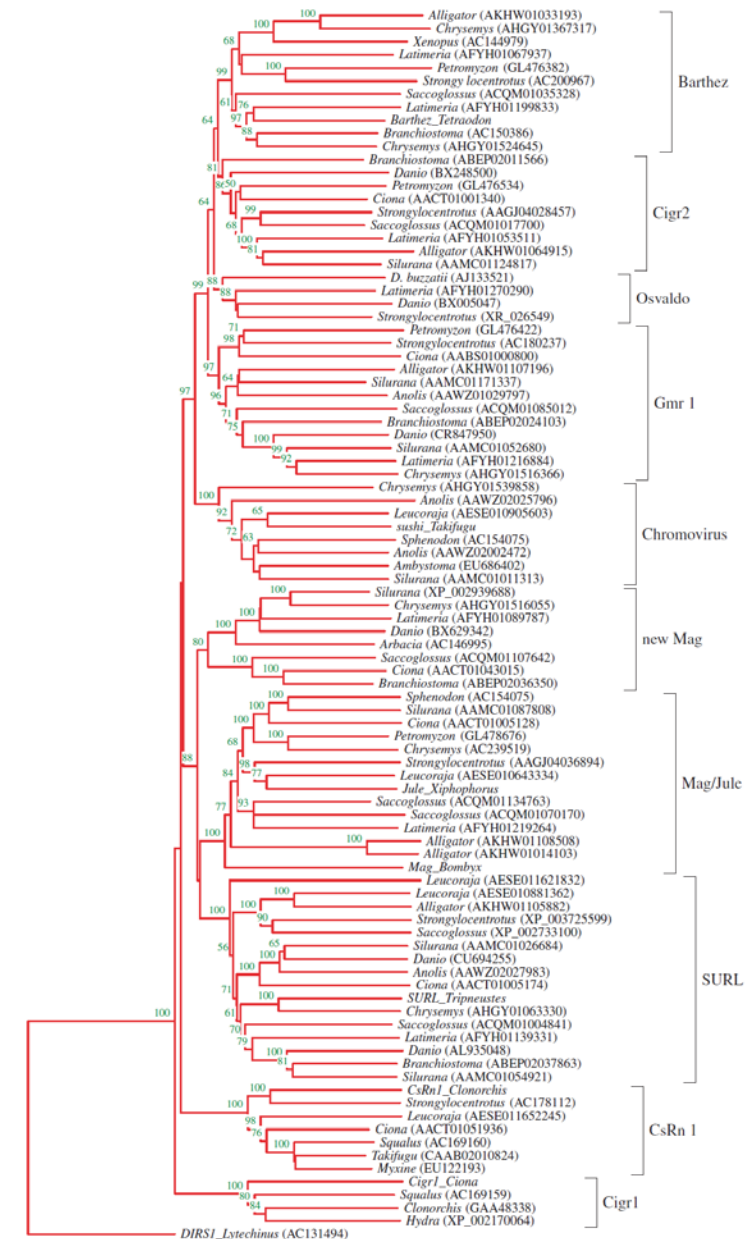
DOMESTICIRANI GENI SO NASTALI IZ OSTANKOV RETROELEMENTOV

Metaviridae clade	Sauropsida				Mammals		Ancestral states			
	Lepidosauria	Turtles	Archosauria (crocodiles)	Archosauria (birds)	Prototheria	Theria	Sauropsida	Synapsida	Amniota	Tetrapoda
Chromovirus	+	*	*	Δ	Δ	Δ	+	+	+	+
Cigr2	Δ	Δ	+	Δ	Δ	Δ	+	+	+	+
Barthez	Δ	+	+	Δ	Δ	Δ	+	+	+	+
Gmr1	+	+	+	Δ	Δ	Δ	+	+	+	+
new Mag	Δ	+	Δ	Δ	Δ	Δ	+	+	+	+
SURL	+	+	+	Δ	Δ	Δ	+	+	+	+
Mag	+	+	+	Δ	Δ	Δ	+	+	+	+

+: full length elements present, Δ: elements absent, *: corrupted elements (retroelement remains)

Filogenetska analiza Metavirid pri kopenskih vretenčarjih: aktivni metaviridni retroelementi (progenitorji domesticiranih genov) so še vedno prisotni v genomih različnih plazilcev.

Domesticirani geni niso povzročili izginotja/utišanja aktivnih skupin metavirid pri sesalcih (nikoli niso bili prisotni v genomih sesalcev !!).



DOMESTICIRANI GENI IN NASTANEK RAZLIČNIH FENOTIPSKIH NOVOSTI

Gag, integrase, and protease domains are highly versatile **protein–protein interaction modules** that can readily interact with novel targets.

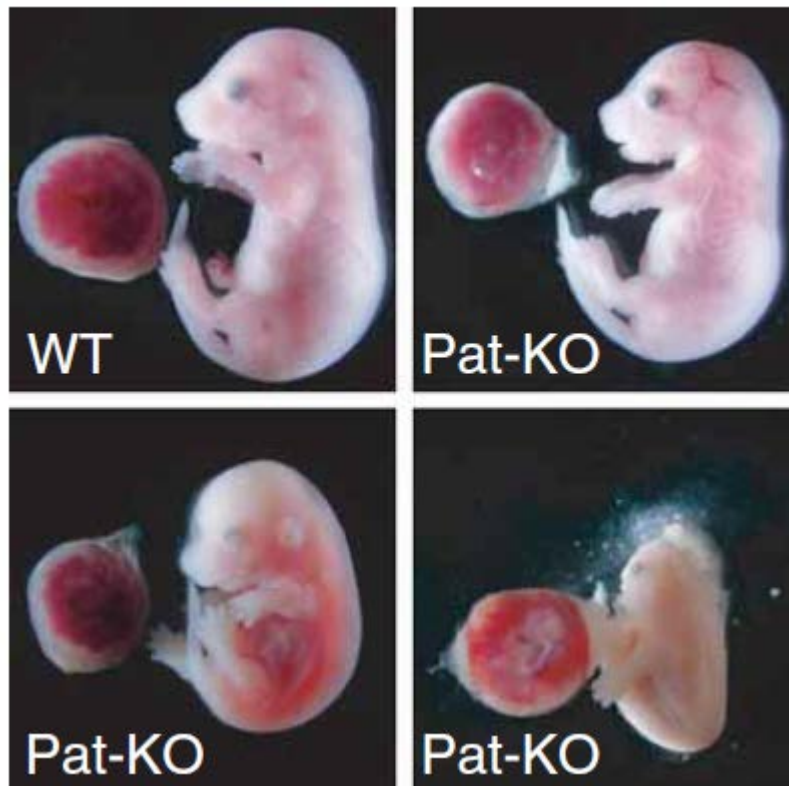
DGs may evolve new functional roles through **adaptive evolution of encoded proteins and/or by developing new spatial or temporal expression patterns.**

Some DGs (e.g., LDOC1) are involved in the **gradual growth of CNS interaction networks** in the particularly active regions of brain (neocortex) - not only during the evolution of placentals, but in very recent times, that is, after the split of Homo and chimpanzee lineages.

Neogen	Ancestralna domena	Ancestralni TE	Funkcija
PEG10	<i>gag</i>	Sushi type LTR retrotransposon	apoptosis, cell proliferation, placenta formation
RTL1	<i>gag</i>	Sushi type LTR retrotransposon	feto-maternal interface, development of placenta
LDOC1	<i>gag</i>	Sushi type LTR retrotransposon	inhibition of NF-kappaB mediated response, potential tumor suppressor
MOAP1	<i>gag</i>	Barthez lineage LTR retrotransposon	BAX binding protein, modulator of apoptosis
GIN1	<i>integraza</i>	Gmr1 clade LTR retrotransposon	unknown
ASPRV1 (SASPase)	<i>proteaza</i>	Cigr2 clade LTR retrotransposon	expressed in epidermis, involved in prevention of wrinkle formation

RTL1 GEN: ESENCIALEN GEN, KLJUČEN ZA RAZVOJ PLACENTE

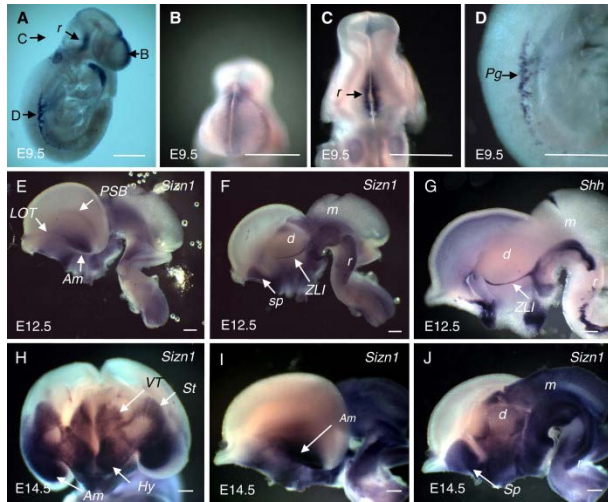
- plays an **essential role** in the **development of the placenta and feto-maternal interface**
- paternally expressed, has an antisense transcript which is maternally expressed



Role of retrotransposon-derived imprinted gene, *Rtl1*, in the feto-maternal interface of mouse placenta

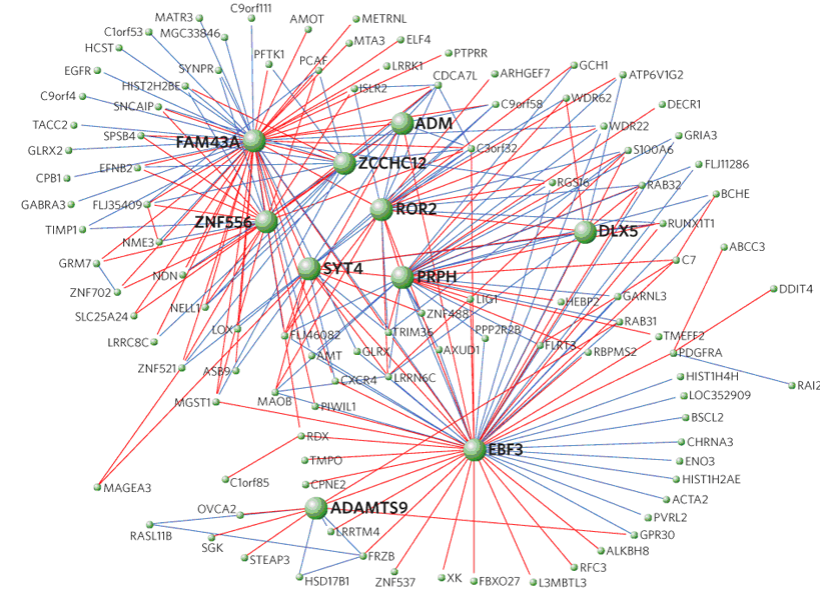
Yoichi Sekita¹, Hiroataka Wagatsuma², Kenji Nakamura³, Ryuichi Ono¹, Masayo Kagami⁴, Noriko Wakisaka^{1,5}, Toshiaki Hino³, Rika Suzuki-Migishima³, Takashi Kohda¹, Atsuo Ogura⁶, Tsutomu Ogata⁴, Minesuke Yokoyama^{3,7}, Tomoko Kaneko-Ishino⁵ & Fumitoshi Ishino¹

ZCCHC12/Sizn1 GEN IN RAZVOJ MOŽGANOV: REGIONALIZACIJA KORTEKSA

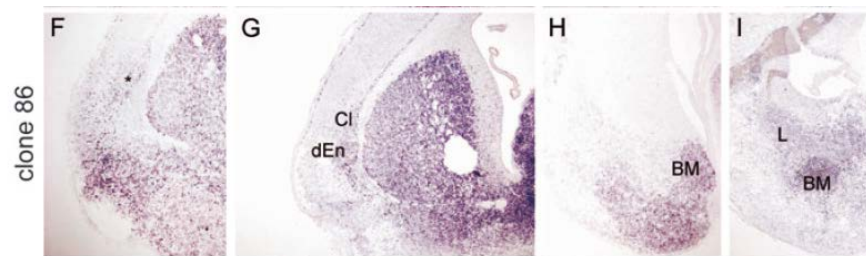
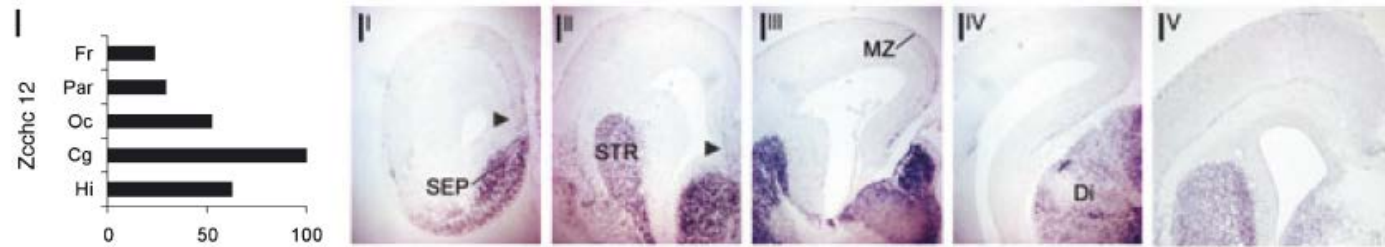


Cho et al., 2011

Genes that may be involved in the cortical regionalization during late neurogenesis in mouse: Differential gene expression across the embryonic cerebral cortex is assumed to play a role in the subdivision of the cortex into distinct areas with specific morphology, physiology and function.



Konopka et al. Nature 2009



Muehlfriedel et al. 2007

MEHANIZMI NEOFUNKCIONALIZACIJE DOMESTICIRANIH GENOV

- V **tranzicijski fazi od ostankov retroelementov do prvotnih DG** so bile potrebne **številne nukleotidne spremembe za neofunkcionalizacijo**.
- Eden od ključnih korakov v procesu neofunkcionalizacije je bila **eksonizacija proteinskih domen retroelementov (gag, proteaza in integraza), ki je privedla do nastanka „ready-to-use“ modulov**.
- Ostanki retroelementov v sesalskih genomih bi se brez promotorjev spremenili v pseudogene. Kot **funkcionalni geni so lahko preživeli le z pridobitvijo novih promotorskih zaporedij**.
- **Novi geni morajo pridobiti promotorske in ostale strukturne elemente, ki regulirajo njihovo izražanje tako, da se bodo izražali v zadostni količini in v primernih tkivih oz. celicah.**

Neaktivno zaporedje (ostanki Metavirid)

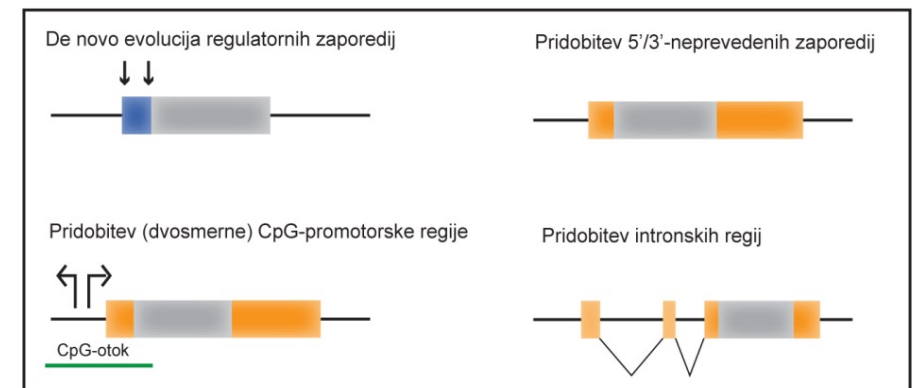


Nukleotidne zamenjave - proces eksonizacije

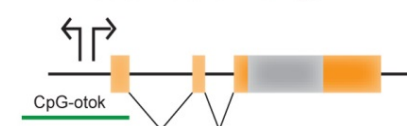
Eksonizirano zaporedje



Mehanizmi, udeleženi pri pridobivanju promotorskih in regulatornih zaporedij



Neofunkcionaliziran gen



Proces molekularne domestikacije je bil počasen: trajal naj bi 90 – 100 milijonov let.

REGULATORNO OŽIČENJE DOMESTICIRANIH GENOV

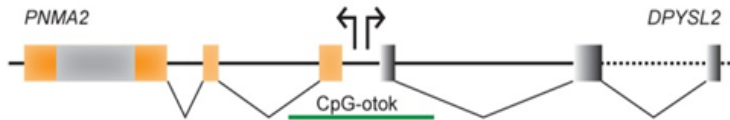
A) De novo pridobitev promotorske regije



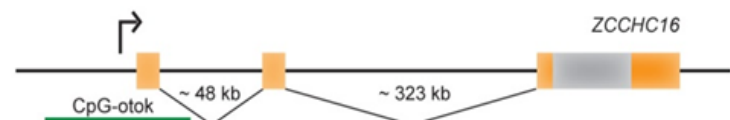
B) CpG-promotor



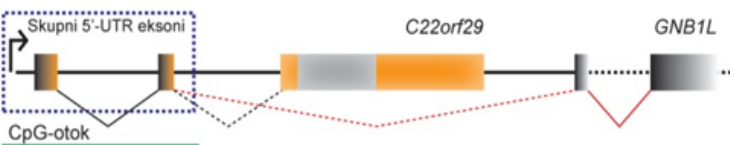
C) Dvosmerni CpG-promotor



D) Pridobitev oddaljenega CpG-promotorja (evolucija 5'-UTR)



E) Skupni promotor



ŠTEVILNI VIRI PROMOTORJEV

Dokazali smo, da so **domesticirani geni** svoje regulatorne in promotorske regije pridobili *de novo*.

Med procesom molekularne domestikacije je bila *de novo* pridobitev regulatornih regij ključna za preživetje domesticiranih genov.

Rezultati naše študije so omogočili vpogled v **nastanek in evolucijo de novo pridobljenih regulatornih regij** (promotorjev, 5' in 3'UTR) pri različnih sesalskih domesticiranih genih.

Pridobitev in evolucija regulatornih zaporedij domesticiranih genov sta igrali pomembno vlogo pri **adaptivni evoluciji** teh genov ter sta pomembni za pojasnitev **tkivno, vrstno in časovno specifičnih vzorcev izražanja** teh genov.

cis-REGULATORNA EVOLUCIJA PRI DOMESTICIRANIH GENIH

Chromovirus-related RDDGs

LDOC1:

Homo: AP-1, ATF-2, AP-2gamma, FOXL1, Egr-4, COMP1, AP-2beta, c-Jun, AP-2alpha, AP-2alphaA
Mus: Elk-1, Meis-1, Meis-1b, HOXA9, HOXA9B, Nkx5-1, SEF-1(1), p300, POU3F2, Nkx6-1

LDOC1L:

Homo: NF-1, Tal-1, CUTL1, Tal-1beta, E47, NRF-2, COMP1, STAT3, ITF-2
Mus: RFX1, PPAR-gamma1, PPAR-gamma2, HOXA9, HOXA9B, Meis-1, Meis-1a, c-Myb, Meis-1b, POU3F2

RGAG4:

Homo: E2F-4, E2F-3a, E2F-5, E2F-2, Egr-4, MZF-1, E2F, E2F-1, COMP1, Pax-4a
Mus: XBP-1, Cdc5, IRF-1, PPAR-gamma1, PPAR-gamma2, HTF, CUTL1, NCX/Ncx, HEN1, GR

FAM127A:

Homo: Elk-1, Nkx5-1, c-Ets-1
Mus: -

FAM127C:

Homo: Elk-1, Nkx5-1, c-Ets-1
Mus: -

PEG10:

Homo: C/EBPbeta
Mus: -

ZCCHC5:

Homo: Sox5, PPAR-gamma1, Olf-1, FOXO1a, PPAR-gamma2, HNF-3beta, Arnt, FOXO1
Mus: POU3F2, POU3F2 (N-Oct-5a), POU3F2 (N-Oct-5b), POU2F1, POU2F1a, Bach1, Pax-6, Brachyury, Evi-1, Roaz

ZCCHC16:

Homo: Sp1, RREB-1, NF-YA, HNF-3beta, NF-YB, CBF-A, CBF-B, CP1A, NF-Y, CBF(2)
Mus: PPAR-gamma1, PPAR-gamma2, C/EBPalpha, LCR-F1, RREB-1, HNF-4alpha1, HNF-4alpha2, FOXD1, NF-E2, NF-E2 p45

RTL1:

Homo: AML1a, Pax-5, Olf-1, MyoD, E4BP4, C/EBPalpha, FOXJ2 (long isoform), FOXJ2
Mus: FOXC1, ARP-1, HFH-1, RP58, STAT1, STAT1alpha, STAT1beta, STAT2, STAT3, STAT4

PNMA family

PNMA1:

Homo: E2F-3a, E2F-4, E2F-5, Brachyury, HSF1 (long), E2F-2, E2F-1, E2F, HSF1short, ATF
Mus: ZIC2/Zic2, Roaz, Nkx3-1v1, Nkx3-1v2, Nkx3-1v3, Nkx3-1v4, Nkx3-1, Msx-1, Zic1, RFX1

PNMA2:

Homo: AML1a, Pax-5, MyoD, Lmo2, AP-4, GATA-1, Egr-4, FOXL1, HEN1
Mus: NRSF form1, NRSF form2, GATA-1, ITF-2, Tal-1beta, HSF1 (long), HSF1 (short), YY1, MyoD, C/EBPalpha

PNMA3:

Homo: E2F-3a, E2F-4, E2F-5, SREBP-1c, E2F-2, SREBP-1b, E2F-1, E2F, SREBP-1a, HOXA5
Mus: ATF2, CRE-BP1, ATF, Ik-3, c-Jun, RP58, NF-kappaB1, Roaz, ISGF-3, p53

MOAP1:

Homo: AML1a, ATF-2, NF-kappaB, FOXL1, AREB6, IRF-2, NF-kappaB2, Meis-1a, NF-kappaB1, RSRFC4
Mus: Evi-1, ZID, c-Myb, STAT3, POU3F1, FOXJ2, FOXJ2 (long isoform), Nkx3-1, Nkx3-1 v1, Nkx3-1 v2

PNMA5:

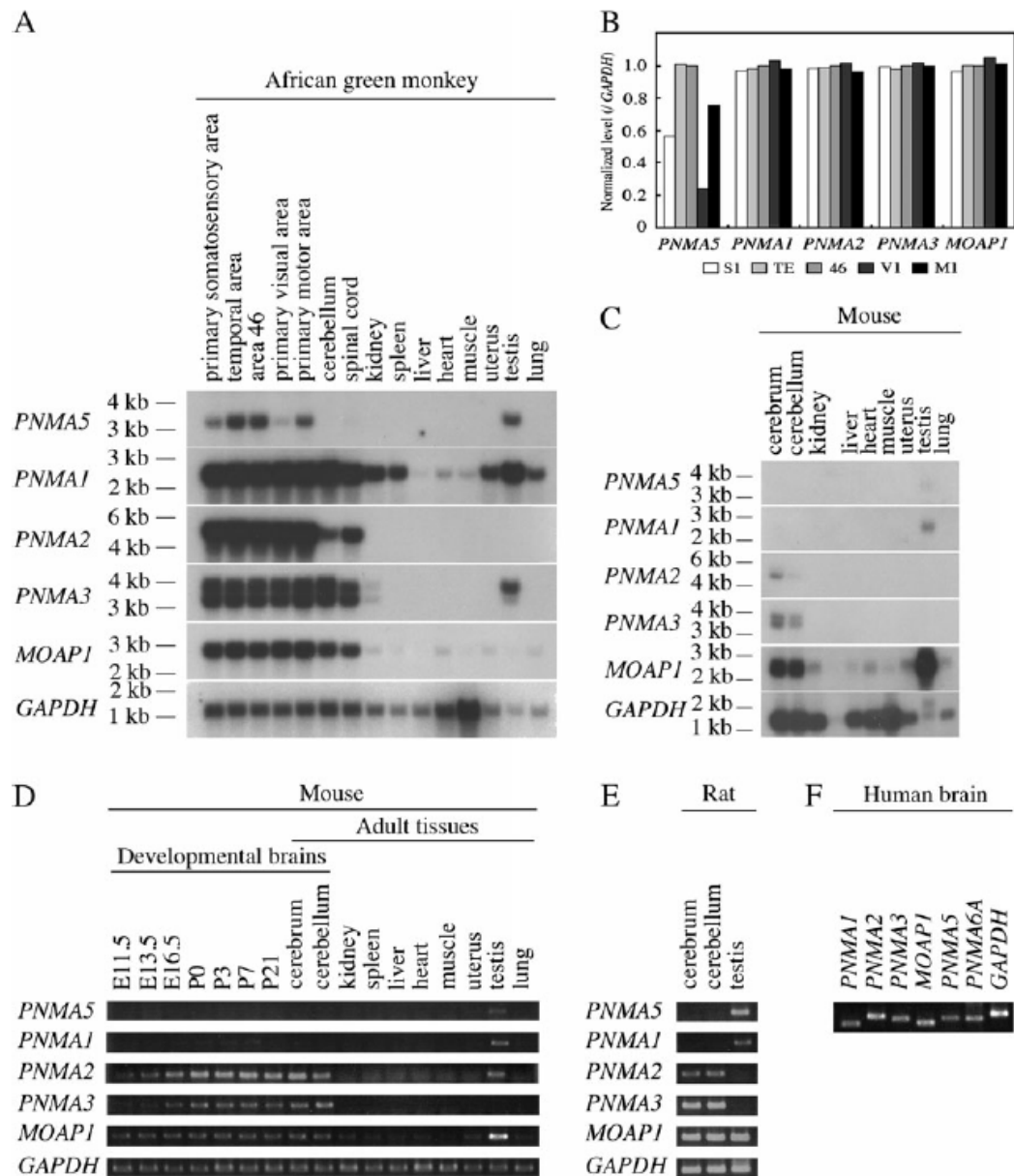
Homo: COUP-TF1, LHX3b/Lhx3b, C/EBPbeta, AML1a, HNF-4alpha2, HNF-4alpha1, GATA-2, COUP-TF, CP2, LHX3a/Lhx3a
Mus: Nkx6-1, Evi-1, HSF1 short, HSF1 (long), Nkx2-5, NF-kappaB1, Pax-2, Pax-2a, SREBP-1a, SREBP-1b

Z analizo vezavnih mest za transkripcijske faktorje v promotorjih domesticiranih genov smo dokazali velike razlike med geni oz. med človeškimi in mišjimi ortolognimi geni.

POSLEDICE *cis*-REGULATORNE EVOLUCIJE: NOVI VZORCI V IZRAŽANJU GENOV

Tkivno, vrstno, časovno in razvojno specifični vzorci v izražanju domesticiranih genov.

Analiza izražanja genov iz PNMA družine v različnih tkivih z hibridizacijo odtisa northern in z RT_PCR (podatki so iz Takaji et al. (2009)).



POMEN DOSEŽKA

Pojasnili smo kako, kdaj in zakaj so iz „genomskih odpadkov“ nastali novi geni (mnogi so esencialni !) z novo biološko vlogo in velikim fenotipskim učinkom (npr. razvoj placente in povečanje velikosti/kompleksnosti možganov)