Identification of a Novel *PNMA-MS1* Gene in Marsupials Suggests the LTR Retrotransposon-Derived *PNMA* Genes Evolved Differently in Marsupials and Eutherians

Sawa Iwasaki^{1,2}, Shunsuke Suzuki^{3,4}, Matthew Pelekanos³, Helen Clark³, Ryuichi Ono¹, Geoff Shaw³, Marilyn B. Renfree³, Tomoko Kaneko-Ishino², and Fumitoshi Ishino^{1,5,*}

Department of Epigenetics, Medical Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan¹; School of Health Sciences, Tokai University, Bohseidai, Isehara, Kanagawa 259-1193, Japan²; Department of Zoology, The University of Melbourne, Melbourne, Victoria 3010, Australia³; Epigenomics Division, Faculty of Agriculture, Frontier Agriscience and Technology Center, Shinshu University, 8304 Minami-minowa, Kami-ina, Nagano 399-4598, Japan⁴ and Global Center of Excellence Program for International Research Center for Molecular Science in Tooth and Bone Diseases, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan⁵

*To whom correspondence should be addressed. Tel. +81-3-5803-4862. Fax. +81-3-5803-4863. Email: fishino.epgn@mri.tmd.ac.jp

Edited by Dr Toshihiko Shiroishi (Received 5 November 2012; accepted 24 April 2013)

Abstract

Two major gene families derived from Ty3/Gypsy long terminal repeat (LTR) retrotransposons were recently identified in mammals. The sushi-ichi retrotransposon homologue (*SIRH*) family comprises 12 genes: 11 in eutherians including *Peg10* and *Peg11/Rtl1* that have essential roles in the eutherian placenta and 1 that is marsupial specific. Fifteen and 12 genes were reported in the second gene family, para-neoplastic antigen MA (*PNMA*), in humans and mice, respectively, although their biological functions and evolutionary history remain largely unknown. Here, we identified two novel candidate *PNMA* genes, *PNMA-MS1* and *-MS2* in marsupials. Like all eutherian-specific *PNMA* genes, they exhibit the highest homology to a Gypsy12_DR (DR, *Danio rerio*) Gag protein. *PNMA-MS1* is conserved in both Australian and South American marsupial species, the tammar wallaby and grey short-tailed opossum. However, no *PNMA-MS1* orthologue was found in eutherians, monotremes or non-mammalian vertebrates. *PNMA-MS1* was expressed in the ovary, mammary gland and brain during development and growth in the tammar, suggesting that *PNMA-MS1* may have acquired a marsupial specific function. However, *PNMA-MS2* seems to be a pseudogene. The absence of marsupial orthologues of eutherian *PNMA* genes suggests that the retrotransposition events of the Gypsy12_DR-related retrotransposons that gave rise to the *PNMA* family occurred after the divergence of marsupials and eutherians.

Key words: LTR retrotransposons; PNMA family; marsupial-specific genes; mammalian evolution

1. Introduction

Approximately 40-50% of the mammalian genome is derived from transposable elements, such as retrotransposons and DNA transposons.¹⁻⁵ The Ty3/Gypsy long terminal repeat (LTR) retrotransposons have been detected in various eukaryotic organisms including fungi, plants, insects, tunicates and echinoderms as well as in several vertebrates, such as fish, amphibians and reptiles, but not in mammals and birds.⁶ However, discrete regions within these elements have acquired new functions as novel endogenous genes and are highly conserved in marsupials and eutherians.^{7–12} Two major gene families derived from

[©] The Author 2013. Published by Oxford University Press on behalf of Kazusa DNA Research Institute.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/ licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com.

the Ty3/Gypsy LTR retrotransposons are the sushi-ichi retrotransposon homologue (SIRH) family (also called the MART or SUSHI family) comprising 12 genes encoding a Gag-like protein, each of which has 20-30% similarity to the sushi-ichi retrotransposon Gag in pufferfish $(Takifugu \ rubripes)$,⁹⁻¹¹ and the para-neoplastic antigen MA (PNMA) family also encoding the Gag-like protein homologous to the Gypsy12 DR retrotransposon Gag in zebrafish (DR, Danio rerio)¹³ comprising 15 and 11 genes in humans and mice, respectively. It should be noted that the homology between these two LTR retrotransposons is only 6.5% and 13.6% along with the entire Gag and Pol regions, respectively. *PEG10/SIRH1* is a therian-specific gene, and *PEG11/* SIHR2 and the remaining SIRH3-11 seem eutherian specific, while SIRH12 was derived from a marsupialspecific retrotransposition event. We previously demonstrated that Peg10/Sirh1 and Peg11/Sirh2 are essential for placental formation and function in mice.^{7,8} Most PNMA genes are expressed in the brains of macagues and mice and their functions remain unknown.¹⁴ PNMA1-3 were first identified as genes encoding neuronal auto-antigens using sera from patients with para-neoplastic neurological syndromes.¹⁵ Schüller et al.¹⁶ and Campillos et al.¹³ performed genome-wide analyses and identified additional 12 family genes in humans among which PNMA6 has no mouse orthologue. No Gypsy12 DR Gag-derived sequences were reported in birds,¹³ and thus, it is probable that the PNMA genes are also mammal specific. However, the search has been limited in several eutherian species and the existence of marsupial orthologues and/or marsupial-specific PNMA genes remained unknown.

Here, we conducted comprehensive *in silico* screening for the *PNMA* genes using the whole-genome shotgun (WGS) sequences of the grey short-tailed opossum⁴ and the tammar wallaby⁵ and identified a novel *PNMA*-*MS1* gene as the first marsupial-specific *PNMA* gene.

2. Materials and methods

2.1. Animals and tissue collection

Tammar wallabies (*Macropus eugenii*) of Kangaroo Island, South Australia origin, were maintained in The University of Melbourne marsupial breeding colony in grassy outdoor enclosures. Lucerne cubes, grass and water were provided *ad libitum* and supplemented with fresh vegetables. The day of birth of pouch young was designated as d0. When the day of birth was unknown, their age was estimated using the head length.¹⁷ Foetal tissues including the head and body were collected from two foetuses at Day 23 and 26 of gestation, and the yolk sac placenta (YSP) from four foetuses sampled between Day 23 and 26 of gestation.

Tissues including the brain, liver, lung, kidney, ovary and testis were collected from two pouch young aged Day 60-70 after birth. The liver, lung, pancreas, stomach, bladder, heart, kidney, adrenal, spleen and brain (cerebrum and cerebellum) were collected from Day 152 and 162 pouch young. Adult female tissues, including the brain (thalamus, hypothalamus and pituitary), ovary (ovary with active corpus luteum, corpus luteum, ovary with developing follicle and ovary with primary or secondary follicle), endometrium (gravid endometrium and non-gravid endometrium) and mammary gland (sucked gland and non-sucked gland) were also collected from two adults. Grey short-tailed opossums (Monodelphis domestica) were purchased from a breeding colony in the Department of Physiology at the University of Melbourne. The brain, liver, spleen, pituitary and ovary were collected from five adult opossums. Experimental procedures conformed to the Australian National Health and Medical Research Council guidelines¹⁸ and were approved by the Animal Experimentation Ethics Committees of the University of Melbourne.

2.2. Reverse transcriptase polymerase chain reaction

Genomic DNA and total RNA from tissues were prepared by using TRIZOL (Invitrogen), as described in the manufacturer's protocol. cDNA was synthesized from 1 µg of total RNA using Superscript III reverse transcriptase (Invitrogen) with an oligo dT primer. Polymerase chain reaction (PCR) amplification for gene expression profiles were carried out using 10–100 ng of cDNA in a 25-µl reaction mixture containing $1 \times ExTag$ buffer, 2.5 mM deoxynucleotide triphosphasteses, 10 pmol primers and 1.25 U ExTag HS (TaKaRa) and were subjected to 30-35 PCR cycles; 96°C for 15 s, 60-65°C for 30 s and 72°C for 15-120 s depending on the length of PCR products at the ratio of 1 min/kb. PCR products were visualized by agarose gel electrophoresis with ethidium bromide staining. The primers used for the expression profiles were as follows: PNMAMS1-F1 (5'-AAC ATG GTG GAG GAG TCT GGA T-3'), PNMAMS1-R1 (5'-CAA CGG TAA GGT GAC CTC TTG G-3'), wGAPDH-F1 (5'-AGA AAG TGG TGA AGC AGG CAT-3'), wGAPDH-R1 (5'-TGG AGG ACA TGT AGA CCA TGA G-3'), wGAPDH-F2 (5'-CCT ACT CCA ATG TAT CTG TGT-3'), wGAPDH-R2 (5'-GGT GGA ACT CCT TTT TTG ACT G-3'), LAMA3-F1 (5'-ACT CTG CAA AGA TCA GCA CAC C-3'), LAMA3-R1 (5'-CTC CTG CCT TCA GCA AGA AGA T-3'); PNMAMS2-F1 (5'-GGC TAA TGG AAA GTC ATA AGA AAG C-3'), PNMAMS2-R1 (5'-GAT TCC TTG ATA CAA ATG GTT GTC C-3'), PNMAMS2-F2 (5'-TTG ATG CAT TGT CTG AAA CCA G-3'), PNMAMS2-R2 (5'-ATC TAT CAA CCA AGC GCC AAC T-3'); oOAZ1-F1 (5'-ATA AAC CCA GCA CCA CCG TCC ACG-3'), oOAZ1-R1 (5'-GGT CTC ACA ATC TCA AAG CCC AAA AAG-3').

2.3. 5'- and 3'-RACE

Rapid amplification of cDNA ends (RACE) reactions were performed with the tammar liver using the RNA SMARTER RACE cDNA Amplification kit (Clontech) according to the manufacturer's recommendations. The 5'- and 3'-RACE fragments were generated with the following gene-specific primers: *PNMAMS1*-5'RACE-GSP1 (5'-TGC GTA TGG AGG GGA GAG TGA GCA AG-3') and *PNMAMS1*-3'RACE-GSP1 (5'-GAC TGT GCC ATC GGG AGA AGG TGA AC-3'), and nested PCR was performed with following primers: *PNMAMS1*-5'RACE-GSP2 (5'-CAG ACA AGG TGG GGT CTG TCT CTT C-3') and *PNMAMS1*-3'-RACE-GSP2 (5'-TTC CTG TGA AGG TCT CCC TCT C-3'), respectively.

2.4. Detection and prediction of the Gypsy12_DR Gag-derived genes

For the detection of PNMA family genes, we performed TBLASTN searches (e-value <1.0E-9) using the NCBI server (http://blast.ncbi.nlm.nih.gov/Blast. cgi) against eutherian reference genomic sequences and marsupial WGS sequences using the Gypsy12-I_DR Gag protein sequence from Repbase (http:// www.girinst.org/) as a query. After TBLASTN searches, sequences that encoded open reading frames (ORFs) with >100 aa (amino acids) were selected for the next analysis. Secondary screening was performed with all the sequences that were selected by first screening as a query. In addition, only sequences encoding proteins with > 100 aa were considered to be candidate PNMA family genes and those with <100 aa were considered as PNMA pseudogenes. Genome resources used were: Homo sapiens (GRCh37.p5), Mus musculus (MGSCv37), M. eugenii (Meug_1.1) and M. domestica (MonDom5). ORF prediction was performed using an ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf. html).

2.5. Multiple alignment and phylogenetic tree

PNMA family genes and retrotransposons from Repbase were aligned using the MEGA 5.0 (Molecular Evolutionary Genetics Analysis). Phylogenetic tree analysis was also performed using the MEGA 5.0. The tree was inferred using the neighbour-joining method with the bootstrap test (1000 replicates). The evolutionary distances were computed using the *p*-distance method and are in the units of the number of amino acid differences per site. All ambiguous positions were removed for each sequence pair.

2.6. Comparative genomic analysis

For comparison of the marsupial *PNMA-MS1* and *-MS2* genomic regions with the corresponding regions in eutherian species, we extracted the following sequences from Ensembl (http://www.ensembl.org);

PNMA-MS1; M. eugenii (GeneScaffold: Meug_1.0: 503: 24720-39177), M. domestica (Chromosome: MonDom5: 3: 260 858 116-260 881 268), H. sapiens (Chromosome: GRCh37: 18: 21 343 369-21 355 887), M. musculus (Chromosome: GRCm37: 18: 12 572 248-12 578 437), Ornithorhynchus ana*tinus* (Chromosome: OANA5: 7: 17 676 195–17 684 806), Gallus gallus (Chromosome: WASHUC2: 2: 106 303 674–106 308 473), *Xenopus tropicalis* (Scaffold: |GI4.1: 84: 2861161-2864246), T. rubripes (Scaffold: FUGU4: 285: 98142-99106) and for PNMA-MS2; M. domestica (Chromosome: MonDom5: 1: 416106001-416736671), Tasmanian devil Sarcophilus harrisii (Scaffold: DEVIL_7.0: GL834637.1: 1-311 016), *H. sapiens* (Chromosome: GRCh37: 9: 125 122 856-125 594 315), M. musculus (Chromosome: GRCm37: 2: 36 078 175-37 218 455), O. anatinus (Chromosome: OANA5 : Ultra70 : 222 463 – 282 938), G. gallus (Chromosome: WASHUC2: 17: 9467 383–9508331), X. tropicalis (Scaffold: |GI_4.2: GL173356.1: 228 227-292 347) and T. rubripes (Scaffold: FUGU4: scaffold 49: 144 872-154 417).

Alignments were obtained using the VISTAWeb server (http://genome.lbl.gov/vista/). *PNMA-MS1* syntenic regions of several species identified above were aligned using the default setting (>70% identity and >100 bp in length) of mVISTA using the LAGAN global multiple alignment option.

3. Results

3.1. Novel candidate PNMA genes in humans and mice

We validated our approach to search for candidate *PNMA* genes in marsupials by performing TBLASTN analysis against human and mouse reference genomic sequences using the Gypsy12_DR Gag protein as a query. With a cut-off *e*-value of <1.0E-9, this screening resulted in 19 and 15 candidates in the human and mouse genomes, respectively. In humans, 15 of the 19 were known *PNMA* genes and the remaining 4 were novel putative *PNMA* genes, *PNMA7/LOC649201, PNMA8/LOC649238, PNMA9/LOC100* 128960 and *PNMA16* (Table 1, Humans). In mice, 12 of the 15 were known, with two novel putative *PNMA* genes, *Pnma7/Gm7028* and *Pnma9/Gm6858*, and one pseudogene, *Gm 1832215* identified (Table 1, Mouse).

The putative human PNMA7-9 genes were located near PNMA6A-D cluster on Chromosome Xq28. There is a sequence gap between PNMA6A-B and 6C-D, so additional PNMA genes may exist in this region (Supplementary Fig. S1). The high homology (47-57%) between the putative amino acid sequences of the PNMA7-9 and PNMA6A-D genes suggested that they share a common ancestor and evolved by gene duplication (see Fig. 3). The putative murine Pnma7, 8 and Table 1. Candidate list for PNMA family genes in two eutherian and two marsupial species

	PNMA number	Gene name	Accession number	Location	
Human	hsPNMA1	PNMA1	NM_006029.4	chr.14:74178486-74181128	
	hsPNMA2	PNMA2	NM_007257.5	chr.8:26362196-26371483	
	hsPNMA3	PNMA3	NM_013364.4	chr.X:152224766-152228827	
	hsPNMA4	PNMA4	NM_022151.4	chr.14:93648541-93651249	
	hsPNMA5	PNMA5	NM_052926.2	chr.X:152157368-152162671	
	hsPNMA6A	PNMA6A	NM_032882.4	chr.X:152338301-152340107	
	hsPNMA6B	PNMA6B	XM_002343859.2	chr.X:152341614-152342813	
	hsPNMA6D	PNMA6D	XM_002343858.2	chr.X:152244152-152246070	
	hsPNMA6C	PNMA6C	NM_001170944.1	chr.X:152240819-152243402	
	hsPNMA7	LOC649201	XP_001127211	chr.X:152584221-152587591	
	hsPNMA8 hsPNMA9	LOC649238 LOC100128960	XM_938309.4 —	chr.X:152662364-152663269 chr.X:152197130-152200901	
	hsPNMA10	ZCCHC12	NM_173798.2	chr.X:117957787-117960931	
	hsPNMA11	ZCCHC18	NM_001143978.1	chr.X:103357107-103360533	
	hsPNMA12	PNMAL1	NM_001103149.1	chr.19:46969748-46974820	
	hsPNMA13	PNMAL2	NM_020709.1	chr.19:46994448-46999169	
	hsPNMA14	CCDC8	NM_032040.3	A_032040.3 chr.19:46913586-46916919	
	hsPNMA15 —		_	chr.19:46931182-46931595	
	hsPNMA16	—	—	chr.19:47036933-47037357	
Mouse	mmPNMA1	Pnma1	NM_027438.3	chr.12:85487081-85489439	
	mmPNMA2	Pnma2	NM_175498.4	chr.14:67530045-67538898	
	mmPNMA3	Pnma3	NM_153169.2	chr.X:70310126-70313530	
	mmPNMA4	Pnma4	NM_001142937.1	chr.12:103978040-103981870	
	mmPNMA5	Pnma5	NM_001100461.3	chr.X:70279327-70282442	
Vlouse	mmPNMA7	Gm7028	NG_005480.3	chr.X;70580917-70581762	
	mmPNMA8	LOC100416956	NG_017874	chr.X:70642228-70644051	
	mmPNMA9	Gm6858	NG_005479.2	chr.X:70295221-70295900	
	mmPNMA10	Zcchc12	NM_028325.3	chr.X:33735899-33739153	
	mmPNMA11	Zcchc18	NM_001035509.1	chr.X:133527694-133531462	
	mmPNMA12	PNMAL1	NM_001007569.1	chr.7:17545144-17547669	
	mmPNMA13	PNMAL2	NM_001099636.2	chr.7:17530031-17532427	
	mmPNMA14 mmPNMA15	CCDC8	NM_001101535.1	chr.7:17579937-17581994 chr.7:17568964-17569311	
		Gm18322	NC_000084.5	chr.18:57308641-57309445	
	mmPNMA pseudo1	UNITOSZZ			
Tammar					
Tammar	mePNMA-MS1	mePNMA-MS1		GeneScaffold_503:27168-31803	
Tammar	mePNMA-MS1 mePNMA pseudo1	mePNMA-MS1			
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2	mePNMA-MS1	_	GeneScaffold_503:27168-31803 Scaffold94060:2914-6134	
Tammar	mePNMA-MS1 mePNMA pseudo1	mePNMA-MS1 	_	GeneScaffold_503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3	mePNMA-MS1 		GeneScaffold_503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4	mePNMA-MS1 		GeneScaffold_503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7	mePNMA-MS1 		GeneScaffold_503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6	mePNMA-MS1 		GeneScaffold 503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold1439:20136-24450 Scaffold391804:1-1797 Scaffold46963:6594-10851	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo8 mePNMA pseudo9	mePNMA-MS1 		GeneScaffold_503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold1439:20136-24450 Scaffold391804:1-1797 Scaffold46963:6594-10851 Scaffold3242:41001-45246	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo8 mePNMA pseudo9 mePNMA pseudo10	mePNMA-MS1 		GeneScaffold_503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold439:20136-24450 Scaffold4391804:1-1797 Scaffold46963:6594-10851 Scaffold3242:41001-45246 Scaffold3242:41001-45246	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo8 mePNMA pseudo9 mePNMA pseudo10 mePNMA pseudo11	mePNMA-MS1 		GeneScaffold_503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold1439:20136-24450 Scaffold391804:1-1797 Scaffold46963:6594-10851 Scaffold3242:41001-45246 Scaffold3242:41001-45246 Scaffold407990:1-2479 Scaffold799:60812-65099	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo8 mePNMA pseudo9 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12	mePNMA-MS1 		GeneScaffold_503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold428007:1-986 Scaffold4391804:1-1797 Scaffold46963:6594-10851 Scaffold3242:41001-45246 Scaffold3242:41001-45246 Scaffold407990:1-2479 Scaffold799:60812-65099 Scaffold492911:1-878	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo8 mePNMA pseudo9 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo13	mePNMA-MS1 		GeneScaffold_503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold1439:20136-24450 Scaffold391804:1-1797 Scaffold46963:6594-10851 Scaffold3242:41001-45246 Scaffold3242:41001-45246 Scaffold407990:1-2479 Scaffold799:60812-65099 Scaffold492911:1-878 Scaffold111753:1-2597	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo1 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo13 mePNMA pseudo14	mePNMA-MS1 		GeneScaffold_503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold1439:20136-24450 Scaffold391804:1-1797 Scaffold46963:6594-10851 Scaffold3242:41001-45246 Scaffold407990:1-2479 Scaffold4799:60812-65099 Scaffold492911:1-878 Scaffold111753:1-2597 Scaffold92201:2820-7152	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo7 mePNMA pseudo10 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo13 mePNMA pseudo14 mePNMA pseudo15	mePNMA-MS1 		GeneScaffold_503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold1439:20136-24450 Scaffold391804:1-1797 Scaffold46963:6594-10851 Scaffold3242:41001-45246 Scaffold407990:1-2479 Scaffold49290:1-2479 Scaffold49291:1-878 Scaffold111753:1-2597 Scaffold92201:2820-7152 Scaffold9921:21555-26131	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo7 mePNMA pseudo10 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo13 mePNMA pseudo14 mePNMA pseudo15 mePNMA pseudo16	mePNMA-MS1 		GeneScaffold_503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold391804:1-1797 Scaffold3242:41001-45246 Scaffold4990:1-2479 Scaffold499:00812-65099 Scaffold111753:1-2597 Scaffold92201:2820-7152 Scaffold9221:21555-26131	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo10 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo13 mePNMA pseudo13 mePNMA pseudo15 mePNMA pseudo16 mePNMA pseudo17	mePNMA-MS1		GeneScaffold 503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold1439:20136-24450 Scaffold391804:1-1797 Scaffold3242:41001-45246 Scaffold3242:41001-45246 Scaffold799:60812-65099 Scaffold492911:1-878 Scaffold111753:1-2597 Scaffold92201:2820-7152 Scaffold921:21555-26131 Scaffold103:4422-8826 Scaffold27816:17763-21999 Scaffold27816:17763-21999	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo10 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo13 mePNMA pseudo14 mePNMA pseudo15 mePNMA pseudo17 mePNMA pseudo17 mePNMA pseudo17	mePNMA-MS1 		GeneScaffold 503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold1439:20136-24450 Scaffold391804:1-1797 Scaffold3242:41001-45246 Scaffold3242:41001-45246 Scaffold407990:1-2479 Scaffold407990:1-2479 Scaffold492911:1-878 Scaffold111753:1-2597 Scaffold92201:2820-7152 Scaffold9221:2155-26131 Scaffold103:4422-8826 Scaffold27816:17763-21999 Scaffold27816:17763-21999 Scaffold134268:3226-6573	
Tammar	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo7 mePNMA pseudo10 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo13 mePNMA pseudo13 mePNMA pseudo15 mePNMA pseudo16 mePNMA pseudo17 mePNMA pseudo17 mePNMA pseudo17	mePNMA-MS1		GeneScaffold 503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold1439:20136-24450 Scaffold391804:1-1797 Scaffold3242:41001-45246 Scaffold3242:41001-45246 Scaffold799:60812-65099 Scaffold492911:1-878 Scaffold111753:1-2597 Scaffold92201:2820-7152 Scaffold921:21555-26131 Scaffold103:4422-8826 Scaffold27816:17763-21999 Scaffold27816:17763-21999	
	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo7 mePNMA pseudo10 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo13 mePNMA pseudo13 mePNMA pseudo15 mePNMA pseudo15 mePNMA pseudo16 mePNMA pseudo17 mePNMA pseudo17 mePNMA pseudo18 mePNMA pseudo19 mdPNMA pseudo19	mePNMA-MS1		GeneScaffold_503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold391804:1-1797 Scaffold3242:41001-45246 Scaffold429090:1-2479 Scaffold499:1:1-878 Scaffold92201:2820-7152 Scaffold921:21555-26131 Scaffold103:4422-8826 Scaffold134268:3226-6573 GeneScaffold_10085:47646-51858	
Tammar Opossum	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo10 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo13 mePNMA pseudo13 mePNMA pseudo14 mePNMA pseudo15 mePNMA pseudo15 mePNMA pseudo16 mePNMA pseudo17 mePNMA pseudo17 mePNMA pseudo19 mdPNMA pseudo19	mePNMA-MS1		GeneScaffold_503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold391804:1-1797 Scaffold3242:41001-45246 Scaffold407990:1-2479 Scaffold4292911:1-878 Scaffold92201:2820-7152 Scaffold921:21555-26131 Scaffold103:4422-8826 Scaffold134268:3226-6573 GeneScaffold_10085:47646-51858	
	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo10 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo13 mePNMA pseudo13 mePNMA pseudo14 mePNMA pseudo15 mePNMA pseudo15 mePNMA pseudo16 mePNMA pseudo17 mePNMA pseudo17 mePNMA pseudo19 mdPNMA-MS1 mdPNMA-MS1 mdPNMA-MS2 mdPNMA pseudo1	mePNMA-MS1		GeneScaffold 503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold1439:20136-24450 Scaffold391804:1-1797 Scaffold3242:41001-45246 Scaffold3242:41001-45246 Scaffold407990:1-2479 Scaffold407990:1-2479 Scaffold492911:1-878 Scaffold111753:1-2597 Scaffold92201:2820-7152 Scaffold92201:2155-26131 Scaffold27816:17763-21999 Scaffold134268:3226-6573 GeneScaffold_10085:47646-51854 chr.3:260874625-260879998 chr.1:416409687-416414127 chr.3:15383915-15388463	
	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo5 mePNMA pseudo7 mePNMA pseudo7 mePNMA pseudo10 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo12 mePNMA pseudo13 mePNMA pseudo14 mePNMA pseudo15 mePNMA pseudo15 mePNMA pseudo17 mePNMA pseudo17 mePNMA pseudo18 mePNMA pseudo19 mdPNMA-MS1 mdPNMA-MS2 mdPNMA pseudo1 mdPNMA pseudo1	mePNMA-MS1		GeneScaffold 503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold385831:1-1432 Scaffold428007:1-986 Scaffold428007:1-986 Scaffold391804:1-1797 Scaffold3242:41001-45246 Scaffold4990:1-2479 Scaffold407990:1-2479 Scaffold492911:1-878 Scaffold49201:2155-26131 Scaffold92201:2820-7152 Scaffold27816:17763-21999 Scaffold134268:3226-6573 GeneScaffold_10085:47646-51854 chr.3:260874625-260879988 chr.1:416409687-416414127 chr.3:239729898-239734143 Scaffold23912-3507	
	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo5 mePNMA pseudo7 mePNMA pseudo7 mePNMA pseudo10 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo13 mePNMA pseudo13 mePNMA pseudo15 mePNMA pseudo15 mePNMA pseudo17 mePNMA pseudo17 mePNMA pseudo18 mePNMA pseudo18 mePNMA pseudo19 mdPNMA-MS1 mdPNMA-MS2 mdPNMA pseudo1 mdPNMA pseudo1	mePNMA-MS1		GeneScaffold 503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold357604:6970-11278 Scaffold428007:1-986 Scaffold428007:1-986 Scaffold391804:1-1797 Scaffold3242:41001-45246 Scaffold42990:1-2479 Scaffold799:60812-65099 Scaffold492911:1-878 Scaffold92201:2820-7152 Scaffold103:442-8826 Scaffold1103:4422-8826 Scaffold134268:3226-6573 GeneScaffold_10085:47646-51853 chr.3:260874625-260879988 chr.1:416409687-416414127 chr.3:15383915-15388463 chr.3:239729898-239734143 chr.1:692451545-692456282 Scaffol2282	
	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo5 mePNMA pseudo7 mePNMA pseudo7 mePNMA pseudo10 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo13 mePNMA pseudo13 mePNMA pseudo15 mePNMA pseudo15 mePNMA pseudo16 mePNMA pseudo17 mePNMA pseudo17 mePNMA pseudo18 mePNMA pseudo18 mePNMA pseudo19 mdPNMA-MS1 mdPNMA-MS2 mdPNMA pseudo1 mdPNMA pseudo1 mdPNMA pseudo1	mePNMA-MS1		GeneScaffold 503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold391804:1-1797 Scaffold3242:41001-45246 Scaffold799:60812-65099 Scaffold492911:1-878 Scaffold92201:2820-7152 Scaffold111753:1-2597 Scaffold9221:2155-26131 Scaffold134268:3226-6573 GeneScaffold_10085:47646-51855 chr.3:260874625-260879988 chr.1:416409687-416414127 chr.3:239729898-239734143 chr.1:451204818-451209222	
	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo6 mePNMA pseudo7 mePNMA pseudo7 mePNMA pseudo10 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo13 mePNMA pseudo13 mePNMA pseudo14 mePNMA pseudo15 mePNMA pseudo16 mePNMA pseudo17 mePNMA pseudo18 mePNMA pseudo18 mePNMA pseudo19 mdPNMA pseudo1 mdPNMA pseudo1 mdPNMA pseudo1 mdPNMA pseudo1 mdPNMA pseudo1 mdPNMA pseudo3 mdPNMA pseudo3 mdPNMA pseudo4 mdPNMA pseudo5	mePNMA-MS1		GeneScaffold 503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold391804:1-1797 Scaffold3242:41001-45246 Scaffold4993:1804:1-1797 Scaffold3242:41001-45246 Scaffold4999:1-2479 Scaffold492911:1-878 Scaffold92201:2820-7152 Scaffold1103:4422-8826 Scaffold1103:4422-8826 Scaffold1103:4422-8826 Scaffold1103:4422-8826 Scaffold1103:4422-8826 Scaffold1103:4422-8826 Scaffold1103:4422-8826 Scaffold1103:4423-8826 Scaffold1103:4423-8826 Scaffold1103:4423-8826 Scaffold1103:4423-8826 Scaffold134268:3226-6573 GeneScaffold_10085:47646-51858 chr.3:260874625-260879998 chr.1:416409687-416414127 chr.3:239729898-239734143 chr.3:239729898-239734143 chr.1:451204818-451209222 chr.1:451204818-451209222 chr.1:451204818-451209222 chr.1:331217342-331	
	mePNMA-MS1 mePNMA pseudo1 mePNMA pseudo2 mePNMA pseudo3 mePNMA pseudo4 mePNMA pseudo5 mePNMA pseudo5 mePNMA pseudo7 mePNMA pseudo7 mePNMA pseudo10 mePNMA pseudo10 mePNMA pseudo11 mePNMA pseudo12 mePNMA pseudo13 mePNMA pseudo13 mePNMA pseudo15 mePNMA pseudo15 mePNMA pseudo16 mePNMA pseudo17 mePNMA pseudo17 mePNMA pseudo18 mePNMA pseudo18 mePNMA pseudo19 mdPNMA-MS1 mdPNMA-MS2 mdPNMA pseudo1 mdPNMA pseudo1 mdPNMA pseudo1	mePNMA-MS1		GeneScaffold 503:27168-31803 Scaffold94060:2914-6134 Scaffold1032:54408-58815 Scaffold385831:1-1432 Scaffold57604:6970-11278 Scaffold428007:1-986 Scaffold391804:1-1797 Scaffold3242:41001-45246 Scaffold799:60812-65099 Scaffold492911:1-878 Scaffold92201:2820-7152 Scaffold111753:1-2597 Scaffold9221:2155-26131 Scaffold134268:3226-6573 GeneScaffold_10085:47646-51855 chr.3:260874625-260879988 chr.1:416409687-416414127 chr.3:239729898-239734143 chr.1:451204818-451209222	

Continued

PNMA number	Gene name	Accession number	Location
mdPNMA pseudo9	_	_	chr.2:271702858-271707121
mdPNMA pseudo10	_	_	chr.5:205686598-205690882
mdPNMA pseudo11	_	—	chr.5:229359947-229364330
mdPNMA pseudo12	_	_	chr.6:103527713-103532099

Candidates in humans, mouse, tammar wallaby and opossum. Newly identified *PNMA* family genes in this study are coloured in grey. 'Pseudo' denotes putative ORFs from sequences detected by TBLASTN, which encode < 100 aa.

9 are all located in the orthologous region on the X chromosome. However, the *PNMA6* cluster is absent from the mouse genome (Supplementary Fig. S1). This region is occupied by the X-linked leucocyte-regulated complex (XIr) gene cluster.

3.2. Identification of novel PNMA genes in marsupials

A comprehensive search of the tammar wallaby (Meug_1.1) and opossum WGS (MonDom5) for PNMA genes was then undertaken using the same method as the human and mouse above. Twenty and 14 hits were returned for TBLASTN searches of the tammar and opossum genomes, respectively (Table 1, Tammar wallaby and Opossum). However, most of the sequences were predicted to be pseudogenes or remnants of the original retrotransposons (<100 aa). Only ORFs predicted to encode >100 aa were considered to be marsupial PNMA candidate genes. One candidate exhibited the highest homology to the Gypsy12_DR Gag protein along with matrix (MA), Nand C-terminal parts of capsid like (N- and C-CA) and cys-cys-his-cys (CCHC) zinc finger domains and had a putative ORF consisting of 456 and 458 aa in the tammar and opossum, respectively (Fig. 1). Therefore, we named it PNMA-MS1 as a novel marsupial-specific PNMA gene. The marsupial PNMA-MS1 gene was located on a syntenic segment in the tammar (Gene scaffold 503:27168-31803) and the opossum (Chr.3: 260 874 625-260 879 998). A second PNMA candidate was identified in the opossum, *PNMA-MS2*, that had a putative ORF encoding 112 aa with high homology only to a central part of the capsid-like domain of the Gypsy12_DR Gag. It was located on Chromosome 1: 416 409 687-416 414 127, where an olfactory receptor (OR) gene cluster exists (see below). The presence of PNMA-MS2 in the tammar was inconclusive due to the incomplete assembly of the corresponding region of the genome.

3.3. Genomic structure of PNMA-MS1 in the tammar wallaby

The full-length sequence of tammar *PNMA-MS1* consisting of 4290 bp was determined by 5'- and 3'-RACE. It has two exons and encodes a putative ORF encoding a

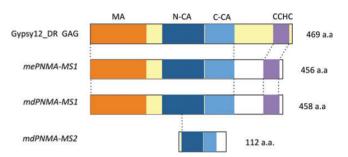


Figure 1. PNMA-MS1 and -MS2 have high homology to the Gypsy12_DR retrotransposon Gag protein. Boxes represent ORFs. Regions with significant similarity to a Gag protein of a Gypsy12_DR retrotransposon are shown in colours. PNMA-MS1 encodes a Gag-like protein including several typical domains and motifs, but lacks a Pol-like protein and LTR sequences attached to either end of the retrotransposons. Although PNMA-MS2 has a possible protein-coding frame corresponding to a central part of the capsid-like domain of the Gypsy12_DR Gag, it seems to be a pseudogene because no expression was confirmed. MA: matrix domain; N-CA: an N-terminal part of capsid-like domain; C-CA: a C-terminal part of capsid-like domain; CHC: a CCHC zinc finger motif for RNA-binding site; RT: reverse transcriptase domain; RNaseH: RNase H domain; INT: integrase domain, me: M. eugenii (tammar wallaby); md: M. domestica (grey short-tailed opossum).



Figure 2. Genomic structure of full-length tammar wallaby *PNMA-MS1*. An arrow represents the direction of *PNMA-MS1* transcription. UTR and ORF are indicated by light blue and dark blue boxes, respectively. There are no supporting data that the promoter of tammar PNMA-MS1 was derived from an LTR sequence of the original retrotransposon. UTR: untranslated region; ORF: open reading frame.

456 aa sequence (Fig. 2). The PNMA-MS1 putative ORF shared 28% similarity at the amino acid level with the Gag protein of Gypsy12_DR retrotransposon (Fig. 3A). Multiple alignment and phylogenetic tree analyses of the putative PNMA-MS1 and -MS2 amino acid sequences are shown in Fig. 3A and B. The Pol protein and LTR regions are absent from PNMA-MS1 and -MS2, suggesting that they no longer have retrotranspositional activity (Fig. 3A). The marsupial PNMA-MS1 and -MS2 protein sequences were grouped together and

Novel Retrotransposon-Derived Genes in Marsupials

A					
GYPSY12_GAG	1 DISQTAQWSTEENINSSRAIVLSNVPLNTSDETIEKVLN-TVKVFGRTQIHGRRG	55	GYPSY12_GAG	327 PSFTQLMKEIREEE-HWVAARVAARVAARVAAR	346
mePNMAMS1 mdPNMAMS1	1 VDEKTLEQWCTALHLDLHRGLLVKGVPHDLTDKDIEKVLHNAISILGKCQVLTKRYEE 1 VDDNILEQWCTVMHLDLHRGLLVKGVPHDLTDKDIENVLHNAIGTLGKCQVLTKRH	58 56	mePNMAMS1 mdPNMAMS1	302 KC-EELMAIVKLYENRLDCGDRSLEDRSLE 307 -GYEMLMETVKLYENRLDCGDRNLQDRNLQ	325 330
mdPNMAMS2	1 0	1	mdPNMAMS2 hsPNMA1	100ID-QWTTIC 319 PNLFQLLVQIREEEAA	107 333
hsPNMA1 mmPNMA1	1 MAMTLLEDWCRGMDVNSORALLVWGIPVNCDETEIEETLQAAMPQVS-YRVLGRMFWR	57 57	mmPNMA1 hsPNMA2	319 PNLFQLLVQIREEE	333 341
hsPNMA2 mmPNMA2	1 MALALLEDWCRIMSVDEQKSLMVTGIPADFEEAEIQEVLQETLKSLGRYRLLGKIFRK 1 MAVALLEEWCKIMGVDVQKSLLVVDIPVDCGEPEIQTVLQEALKCVGSYRLLGKIFQK	58 58	mmPNMA2	322 PTFLOLMKVIREEEEE-EDAYFEDAYFE	343
hsPNMA4 mmPNMA4	1 MTLRLLEDWCRGMDMNPRKALLIAGTSQSCSVAETEEALQAGLAPLGEYRLLGRMFRR 1 MTLRLLEDWCRGMDMNPRKALLVAGIPPTCGVADIEEALQAGLAPLGEHRLLGRMFRR	58 58	hsPNMA4 mmPNMA4	320 PGFLQLLVLIKDYEAAAA	335 344
hsPNMA3 mmPNMA3	1 MPLTLLQDWCRGEHLNTRRCMLILGIPEDCGEDEFEETLQEACRHLGRYRVIGRMFRR	58 58	hsPNMA3 mmPNMA3	321 PGFLALVKLLREEE-EWEATLGATLG	341 342
hsPNMA5	1 WALTLLEDWCKGMDMDPRKALLIVGIPMECSEVEIQDTVKAGLQPLCAYRVLGRMFRR	58	hsPNMA5	316 PNFLELMKLIRDEE-EWENTEAVMKEAVMK	339
mmPNMA5 hsPNMA10	1 MA-SIIAR-VGNSRR	58 18	mmPNMA5 hsPNMA10	207 PNFLELIRMVREEE-DWDDAFIKDAFIK	375 228
mmPNMA10 hsPNMA11	1 MA-SILSR-LGSSRGQNSPL	18 12	mmPNMA10 hsPNMA11	207 PNFLELIRMIREEE-EWEETFINETFINDAFIK	228 228
mmPNMA11	1 MAVGNSR	27	mmPNMA11	208 PDFLELIRMIREEE-DWDETFLRETFLR	229
GYPSY12_GAG	56 DVTGKHLFVLVETRADLDPSTIPPEIGIESEAGPWPVHFVGRLQVQNPAPENDTFQS-	112	GYPSY12_GAG mePNMAMS1	347ENVKLSKAPQVLPMETRMEPP	350 346
mePNMAMS1 mdPNMAMS1	59 TDPTLSVFCKLSEPIDYSKIPSAITVGEN-TWKLVIRPPPEEEE-IER- 57 EDTDLTLSVFCKLSEPIDYSKIPSSLTVGEN-TWKLVSRLPTEEEE-IES	104 104	mdPNMAMS1 mdPNMAMS2	320	343
mdPNMAMS2 hsPNMA1	2 58 EENAKAALLELTGAVDYAAIPREMP-GKGGVWKVLFKPPTSDAE-FLE-	2	hsPNMA1	334KEEE	111 337
mmPNMA1 hsPNMA2	58 EENAKAALLELTGAVDYSLIPREMP-GKGGLWKVVFKPTSDAV-FLE- 59 QENANAVLLELLEDTDVSAIPSEVO-GKGGVWKVIFKTPNQDTE-FLE-	103 104	mmPNMA1 hsPNMA2	334KKEE 342ENES	337 345
mmPNMA2	59 QDNTSVVLVELMEDTDMSVVPSEVQ-GKGGVWKVIFKTPNQDTE-FLQ-	104	mmPNMA2 hsPNMA4	344QESR 336EEEE	347 339
hsPNMA4 mmPNMA4	59 DENRKVALVGLTAETSHALVPKEIP-GKGGIWRVIFKPPDPDNT-FLS- 59 DENKNVALIGLTVETGSALVPKEIP-AKGGVWRVIFKPPDTDSD-FLC-	104 104	mmPNMA4	345AELE	348
hsPNMA3 mmPNMA3	59 EENAQAILLELAQDIDYALLPREIP-GKGGPWEVIVKPRNSDGE-FLN- 59 EENAQAFLVELARDFDYALVPREIE-GKGGPWEVVVKPPHSDDE-FLN-	104 104	hsPNMA3 mmPNMA3	343SERSCYEGLELGPSP	356 357
hsPNMA5 mmPNMA5	59 EDNAKAVFIELADTVNYTTLPSHIP-GKGGSWEVVVKPRNPDDE-FLS-	104	hsPNMA5 mmPNMA5	340NKEKP 376 MTSVKRRRLLWRHSAGEEGQRKESGFWAESEPDEQKPYVRAQESGNERGAWAVSHPNPKE	344 435
hsPNMA10	59 EDEAKAVLIELPEVVDYTMMPTHIP-AEGGAWEVVVKPRSPDDE-FMN- 19AHSMLR-	27	hsPNMA10 mmPNMA10	229RKRPKRSESMVERAVSPVAFQGSPPIV 229PKRPRRAESVMERALSPMAFQSSPPIM	255 255
mmPNMA10 hsPNMA11	19AHSMLR- 13 QQNAAHSMLR- 	27 27	hsPNMA11	229RKRPKRSEPIMERAASPVAFQGAQPIA	255
mmPNMA11	28RSU	33	mmPNMA11	230NKRPRRSETVMERAASPVVFQGSLPIV	256
GYPSY12_GAG	113 KLLTLMQQEGKSMDEVKAILMGEHSPKSDINVDLVDAIGKLVDRCN	158	GYPSY12_GAG mePNMAMS1	351ASVATVISPQSDGPSELQSLKKEVKELSSQ 347HSHLGTKDETSNMPVEKRNAG	380 367
mePNMAMS1 mdPNMAMS1	105 KLKMVLQKEGLSEDDVKH-AYGGE-SEQDESEPEEDS 105 KLKTVLQREGLS-EDVVRYVYGGKTRPGESEQAEY	139 138	mdPNMAMS1	344PKEKKMEQSTSHAGMKTHKKASEEWNSG	371
mdPNMAMS2 hsPNMA1	2 104 RUHLFLAREGWTVQDVAR-VLGFQNPTP-TPGPEMPAEMLNYILDNVI	2 149	mdPNMAMS2 hsPNMA1	111 338EEAEATLLQL	111 347
mmPNMA1	104 RLHLFLAREGWTVQDVAR-VLGFQNPAP-APGPEMPAEMLNYILDNVI	149	mmPNMA1 hsPNMA2	338ERAEAALLQL 346EEP	347 349
hsPNMA2 mmPNMA2	105 RLNLFLEKEGQTVSGMFR-ALGQEGVSP-ATVPCISPELLAHLLGQAMAHAP 105 RLNLFLEKEGQTVAGMFR-ALKHEGVSP-ATPPCTSPEL-LAHLTGQAMVHGQ	154 154	mmPNMA2	347	347
hsPNMA4 mmPNMA4	105 RLNEFLAGEGMTVGELSR-ALGHENGSLDPEQGMIPEMWAPMLAQAL-EAL 105 RLNEFLKGEGMTMGELTR-VLGNRNDPLGL-DPG-IMIPEI-RAPMLAQALNEAL	153 155	hsPNMA4 mmPNMA4	339 348	339 348
hsPNMA3	105 RLNRFLEEERRTVSDMNR-VLGSDTNCS-APRVTISPEFWTWAQTLGAAV	152	hsPNMA3 mmPNMA3	357PARITGVGAVPLPASGNSFDVRPSQGYRRR	386 387
mmPNMA3 hsPNMA5	105 RLNHFLEEERRTVSDMNR-VLGTHSNH-SPTKTTISADFWVWAQTLGAVM 105 RLNYFLKDEGRSMTDVAR-ALGCCS-LPAESLDAEVMPQVRSPPL	152 147	hsPNMA5	358SNIGSEERELFVPAFGSVLEERPYQGSRRR 345SGRGRGASGRQARAEASVSAPQATVQAR	372
mmPNMA5 hsPNMA10	105 KLIYFLRDEGRRIVDVAK-ALGFSTVPTGKIELKNLDQDKPKGL 28 SLGRSLG	147 34	mmPNMA5 hsPNMA10	436 IEAQDSQEFLPVAGNRDTLTKSWGSPDKGTGDMSVAEGQQGQGKAPNFLLARNDPNKQEQ 256 IGSADCNVIEIDDTLDDSDEDVILVESODPPLPSW	495 290
mmPNMA10	28 SLGRSLGG	34	mmPNMA10 hsPNMA11	256 ISSIDCNVIEIDDSPDDSDEDVILVEPEDPPLPSS 256 ISSADCNCNVIEIDDTLDDSDEDVILVVSLYPSLTPT	290 292
hsPNMA11 mmPNMA11	28 SLGRSLGC	34 34	mmPNMA11	250 ISSAUCHCNVIEIDDILDDSQDDSDEDVILVESLIPSLIPS	291
GYPSY12_GAG	159 QASN-DGPSYR-KLRLFSGLKPVP-PGEEEYEIWMEQAAQMISEWQCTEAS	215	GYPSY12_GAG	381 MSHLLNVATATCASECAPQKTSSKNSESV-KRD-KSQPTKLTQQPV	424
mePNMAMS1 mdPNMAMS1	139 QASM-DUF3M, -KIKLESULAPVF-DEGETELINNEQAAQMISENQLERSSAQKIVES 140 QPM	189 194	mePNMAMS1 mdPNMAMS1	368 QFPFCQKKNKGLWGQRSSRPEARTWPVE-QRQRGPG 372 QYHFSQKKGLWGQRANQLEARTWPIE-QRQRGPI	402
mdPNMAMS2	3	11	mdPNMAMS2	572 QINTSQAKUQAKUQAKU	112
hsPNMA1 mmPNMA1	150 QPLV-ESIWYK-RLTLFSGRDI-PGPGEETFDPWLEHTNEVLEEWQVSDVERRRLMESL 150 QPLV-ESIWYK-KLTLFSGKDI-PGPGEETFDSWLEHSNEIIEEWQVSDIERRRLMESL	206 206	hsPNMA1 mmPNMA1	348	349 349
hsPNMA2 mmPNMA2	155 QPLL-PMRYR-KLRVFSGSAV-PAPEESFEVWLEQATEIVKENPYTEAEKKWLAESL 155 RPLL-PVKYC-KMRIFSGSTA-AAPEEPFEVWLEQATEIXKEMPIPEAEKKWVAESL 154 QPAL-QCLKYK-KLRVFSGRES-PEGEEFFGRWIHFHTQVKANQVPDVEYRRLIESL 156 KPTL-QYLRYK-KLSVFSGRDP-PGPGEEFESIMFHTSQVMKTNQVSDVEYRRLIESL	210 210	hsPNMA2 mmPNMA2	350ERDGYG	356
hsPNMA4	154 QPAL-QCLKYK-KLRVFSGRES-PEPGEEEFGRWMFHTTQMIKAWQVPDVEKRRRLLESL	210			
mmPNMA4 hsPNMA3			hsPNMA4	340LOAI	357 345
	156 KPTL-QYLRYK-KLSVFSGRDP-PGPGEEEFESWMFHTSQVMKTWQVSDVEGRRRLIESL 153 QPLL-EQMLYR-ELRVFSGNTI-SIPGALAFDAWLEHTTEMLQMWQVPEGEGRRRLMECL	212 209	hsPNMA4 mmPNMA4 hsPNMA3	340LQAI 349	357
mmPNMA3 hsPNMA5	153 QPLL-EQMLTR-ELRVFSGN11-SIPGALAFDAWLEHTTEMLQMWQVPEGERKRKLMECL 153 OPLL-FOMLYR-FLRVFSGNTT-SIPGLLAFDSWLEHTTEMLOMWQVPEVERRRLMECL	212	mmPNMA4 hsPNMA3 mmPNMA3	340	357 345 350 409 412
mmPNMA3 hsPNMA5 mmPNMA5	153 QPLL-EQMLTR-ELRVFSGN11-SIPGALAFDAWLEHTTEMLQMWQVPEGERKRKLMECL 153 OPLL-FOMLYR-FLRVFSGNTT-SIPGLLAFDSWLEHTTEMLOMWQVPEVERRRLMECL	212 209 209 204 205	mmPNMA4 hsPNMA3 mmPNMA3 hsPNMA5 mmPNMA5	340 AL	357 345 350 409 412 401 551
mmPNMA3 hsPNMA5	133 OPLI-EQMIX NA-ELKVPSON II-SIMGALAPOANLEH I ENLOMIQVPGOSKNIKINEL 133 OPLI-EQMIX PA-ELKVPSON II-SIMGALAPOANLEH I ENLOMIQVPGVSOKKNIKEL 148 ESPK-ESMIX PA-KLKVPSOTA-BSPGETFEDNLEQVTEINFINOVSEV ORRILLES 148 ESICMISTCIX-KLKVPSOFP-ROGEESFETWLEQVTEINFINOVSEV ORRILLES 135 -PIM-ASMADR-HMKLPSGRVV-PAOGEETFENNLQVVGVLPDINMSEESUKRIKHTL 135 -PIM-ASMADR-HMKLPSGRVV-PAOGEETFENNLQVVGVLPDINMSEESUKRIKHTL	212 209 209 204	mmPNMA4 hsPNMA3 mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 mmPNMA10	340 -AL -LQAT 349 -G' -G' 348 RGRGQ -HRG -G' 387 RGRGQ -HRG -G' 373 SFDSSPQ -TIQ -G' 496 I-PHSSVTTKWQDRGECQRLKW -GASMITPQGNPDRSWDTS-GSQDEDGCSELRMPT 291 GAPPLRDRARPQDEVLVIDSPH-NS	357 345 350 409 412 401
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 mmPNMA10 hsPNMA11	133 QPLL-EQMIX AR-ELKYPSONII-SIPGALAPDANLEHIIBMLQMMQVPEGSKKKINBLL 133 QPLL-EQMIX AR-ELKYPSONII-SIPGALAPDANLEHIIBMLQMMQVPEGSKKKINBLL 148 ESPK-ESMMYR-KLIKVPSGTS-BSPGEFFEDNULEQVTEINDINQVSEVE ARRILESL 135 -PIM-ASMADR-NMKLFSGRVP-PAOGEEFFENNULTQVNGVLPDINNSEEFLKRLMKTL 35 -PIM-ASMADR-NMKLFSGRVP-PAOGEEFFENNULTQVNGVLPDINNSEEFLKRLMKTL 35 -PIM-ASMADR-NMKLFSGRVP-PAOGEEFFENNULTQVNGVLPDINNSEEFLKRLMKTL	212 209 209 204 205 90 90 90	mmPNMA4 hsPNMA3 mmPNMA3 hsPNMA5 mmPNMA10 mmPNMA10 hsPNMA10 hsPNMA11	340	357 345 350 409 412 401 551 342 343 344
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 mmPNMA10 hsPNMA11 mmPNMA11	133 QPLL-EQMIX AR-ELAVPSONIT-SIGGLAFDAMLEHTIENLQMMQVPEGESAKAKINAEL 133 QPLL-EQMIX AR-ELAVPSONIT-SIGGLAFOSMLEHTIENLQMMQVPEGESAKAKINAEL 148 ESPK-ESMMYA-KLKVPSGTS-BSPGEETFEDMLEQVTEINDINQVSEVE ARRILESL 135 -PIM-SIMADR-NMKLFSGRVP-PAOGEETFENNLTQVNGVLPDINMSEEFLKRLMKTL 35 -PIM-ASMADR-NMKLFSGRVP-PAOGEETFENNLSQVTGVLPDINMSEEFLKRLMKTL 35 -PLV-VKMAER-INKLFSGRVP-PAOGEETFENNLSQVTGVLPDINMSEEFLKRLMKTL 35 -PLI-ANIAER-NIQSFSGRAE-LGPGEETFENNLSQVTBVLPDINMSEEFLKRLMKTL	212 209 204 205 90 90 90 90	mmPNMA4 hsPNMA3 mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 mmPNMA10 hsPNMA11 mmPNMA11	340	357 345 350 409 412 401 551 342 343 344 343
mmPNMA3 hsPNMA5 hsPNMA10 mmPNMA10 hsPNMA11 mmPNMA11 gYPSY12_GAG mePNMAM51	133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQMOVPEGESKAKKINAEL 133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQMOVPEGESKAKKINAEL 148 ESEKOEN SENKEN SCHLESSEN SENKEN S	212 209 204 205 90 90 90 90 274 248	mmPNMA4 hsPNMA3 mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 mmPNMA10 hsPNMA11 mmPNMA11 GYPSY12_GAG mePNMAMS1	340	357 345 350 409 412 401 551 342 343 344 343 464 443
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 mmPNMA10 hsPNMA11 mmPNMA11 GYPSY12_GAG	133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQMOVPEGESKAKKINAEL 133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQMOVPEGESKAKKINAEL 148 ESEKOEN SENKEN SCHLESSEN SENKEN S	212 209 204 205 90 90 90 90 90	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 hsPNNA5 hsPNNA10 mmPNNA10 hsPNNA11 mmPNMA11 GYPSY12_GAG mePNMAMS1 mdPNNAMS1	340	357 345 350 409 412 401 551 342 343 344 343 464 443 445 112
mmPNMA3 hsPNMA5 hsPNMA10 hsPNMA10 hsPNMA11 mmPNMA11 mmPNMA11 gYPSY12_GAG mePNMAMS1 mdPNMAMS1 mdPNMAMS2 hsPNMAA	133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQMOVPEGESKAKKINAEL 133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQMOVPEGESKAKKINAEL 148 ESEKOEN SENKEN SCHLESSEN SENKEN S	212 209 204 205 90 90 90 90 274 248 253 70 265	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 hsPNNA5 hsPNNA10 mmPNNA10 hsPNNA11 mmPNNA11 GYPSY12_GAG mePNAMS1 mdPNNAMS1 mdPNNAN52 hsPNNA1	340	357 345 350 409 412 401 551 342 343 344 343 443 443 445 112 353
mmPNMA3 hsPNMA5 hsPNMA10 hsPNMA10 hsPNMA10 hsPNMA11 mmPNMA11 mePNMA11 mePNMAM51 mdPNMAM52 hsPNMA1 mmPNMA1 msPNMA2	133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQMOVPEGESKAKKINAEL 133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQMOVPEGESKAKKINAEL 148 ESEKOEN SENKEN SCHLESSEN SENKEN S	212 209 204 205 90 90 90 90 274 248 253 70 265 265 269	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 mmPNNA5 hsPNNA10 mmPNNA10 hsPNNA11 GYPSY12_GAG mmPNNA51 mdPNNAN51 mdPNNAN51 hsPNNA1 mmPNNA1 hsPNNA2	340	357 345 350 409 412 401 551 342 343 344 343 464 443 445 112 353 353 354
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 hsPNMA10 hsPNMA11 mmPNMA11 GYPSY12_GAG mePNNAMS1 mdPNMAMS1 hsPNMA1	133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQMOVPEGESKAKKINAEL 133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQMOVPEGESKAKKINAEL 148 ESEKOEN SENKEN SCHLESSEN SENKEN S	212 209 209 204 205 90 90 90 90 90 274 248 253 70 265 265	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 mmPNNA5 hsPNNA10 mmPNNA10 mmPNNA11 GYPSY12_GAG mePNNAN51 mdPNNAN51 mdPNNAN51 hsPNNA1 mmPNNA1 hsPNNA2 msPNNA4	340	357 345 350 409 412 401 551 342 343 344 343 464 443 445 112 353 353 364 365 351
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 hsPNMA10 hsPNMA10 dsPNMA11 mmPNMA11 mdPNMA851 mdPNMA851 hsPNMA2 hsPNMA2 hsPNMA2 hsPNMA2 hsPNMA4	133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQMOVPEGESKAKKINAEL 133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQMOVPEGESKAKKINAEL 148 ESEKOEN SENKEN SCHLESSEN SENKEN S	212 209 204 205 90 90 90 90 274 248 253 70 265 265 269 269 269 269 269	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 mmPNNA30 mmPNNA30 hsPNNA10 mmPNNA11 mmPNNA11 mdPNNAM51 mdPNNAM51 mdPNNA5 hsPNNA1 mmPNNA1 hsPNNA2 hsPNNA2 hsPNNA2 hsPNNA4	340 -AL -LQAT 349	357 345 350 409 412 401 551 342 343 344 343 445 112 353 353 364 365 351 352
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 hsPNMA11 gYPSY12_GAG mmPNMA11 GYPSY12_GAG mdPNMAMS1 mdPNMAMS1 hsPNMA2 hsPNMA2 hsPNMA2 hsPNMA4 hsPNMA3 mmPNMA3	133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQMOVPEGESKAKKINAEL 133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQMOVPEGESKAKKINAEL 148 ESEKOEN SENKEN SCHLESSEN SENKEN S	212 209 204 205 90 90 90 90 274 248 253 70 265 265 265 269 269 269 269 271 268	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 mmPNNA3 hsPNNA5 mmPNNA3 hsPNNA10 hsPNNA11 GYPSY12_GAG mdPNNAM51 mdPNNAM51 mdPNNA5 hsPNNA2 hsPNNA2 mmPNNA1 hsPNNA2 mmPNNA1 hsPNNA2 mmPNNA1 hsPNNA2 msPNNA3	340 -AL -LQAT 347 -CQAT	357 345 350 409 412 401 551 342 343 344 343 464 443 445 112 353 353 364 353 353 364 355 351 352 453
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 mmPNMA10 hsPNMA11 GYP5Y12_GAG mePNMA451 mdPNMA451 mdPNMA451 mdPNMA451 msPNMA2 mmPNMA2 mmPNMA2 mmPNMA4 mmPNMA3	133 OPLL - EVMLYA-ELKVPSONT 1-510GALAPOANLEHT I BALQMOVPGOESKAKKINBEL 133 OPLL - EVMLYA-ELKVPSONT 1-510GALAPOANLEHT I BALQMOVPGOESKAKKINBEL 148 ESEKOENIYA-ELKVPSOTA 5-PSPGEETEFDULEOVTEIMPINVSEVEVRRELLESL 148 ESEKOENISTEVK-LUKVPSOTA 5-PSPGEETEFDULEOVTEIMPINVSEEVEVRRELLESL 135 - PIM-ASMAGP-NMKLPSORVV-PAOGEETEFNULTOVNGVLPDINMSEEFELKKINKTI 135 - PLV-ASMAGP-NMKLPSORVV-PAOGEETEFNULTOVNGVLPDINMSEEFELKKINKTI 135 - PLV-ASMAGP-NMKLPSORVV-PAOGEETEFNULSOVHEVLPDIMSSEEFELKKINKTI 135 - PLV-VKMAER-NMKLPSORVV-PAOGEETEFNULSOVHEVLPDIMSSEEFELKKINKTI 136 - PLI-AMIAER-MIQSFSORAE-LGØGEETEFNULSOVHEVLPDIMSSEEFELKKINKTI 137 - PLI-AMIAER-NIGSFSORAE-LGØGEETEFENULSOVHEVLPDIMSSEEFELKKINKTI 138 - PLI-AMIAER-NIGSFSORAE-LGØGEETEFENULSOVHEVLPDIMSSEEFELKKINKTI 139 KORAPOLISILIKUSOPSATATEYMAALETAVGTYECOPDIAM-KERHTYODNGTU BAFL	212 209 204 205 90 90 90 90 274 248 253 70 265 265 265 265 269 269 269 269	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 mmPNNA5 hsPNNA10 mmPNNA10 mmPNNA11 GYPSY12_GAG mePNNAM51 mdPNNAM51 mdPNNAM51 hsPNNA1 hsPNNA1 hsPNNA2 hsPNNA3 mmPNNA4 hsPNNA3 hsPNNA3 hsPNNA3 hsPNNA3	340 -AL -LQAT 347 -CQAT	357 345 350 409 412 351 342 343 344 343 464 443 445 353 353 354 355 351 352 450 453 455 605
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 mmPNMA10 hsPNMA11 GYPSY12_GAG mePNMA451 mdPNMA451 mdPNMA451 msPNMA2 hsPNMA2 mmPNMA2 mmPNMA3 mmPNMA3 mmPNMA3 mmPNMA3 mmPNMA3 mmPNMA3 mmPNMA3 mmPNMA5 mmPNMA5 mmPNMA5 mmPNMA5 mmPNMA5	133 OPLL - EVMLYA-ELKVPSONT 1-510GALAPOANLEHT I BALQMOVPGOESKAKKINBEL 133 OPLL - EVMLYA-ELKVPSONT 1-510GALAPOANLEHT I BALQMOVPGOESKAKKINBEL 148 ESEKOENIYA-ELKVPSOTA 5-PSPGEETEFDULEOVTEIMPINVSEVEVRRELLESL 148 ESEKOENISTEVK-LUKVPSOTA 5-PSPGEETEFDULEOVTEIMPINVSEEVEVRRELLESL 135 - PIM-ASMAGP-NMKLPSORVV-PAOGEETEFNULTOVNGVLPDINMSEEFELKKINKTI 135 - PLV-ASMAGP-NMKLPSORVV-PAOGEETEFNULTOVNGVLPDINMSEEFELKKINKTI 135 - PLV-ASMAGP-NMKLPSORVV-PAOGEETEFNULSOVHEVLPDIMSSEEFELKKINKTI 135 - PLV-VKMAER-NMKLPSORVV-PAOGEETEFNULSOVHEVLPDIMSSEEFELKKINKTI 136 - PLI-AMIAER-MIQSFSORAE-LGØGEETEFNULSOVHEVLPDIMSSEEFELKKINKTI 137 - PLI-AMIAER-NIGSFSORAE-LGØGEETEFENULSOVHEVLPDIMSSEEFELKKINKTI 138 - PLI-AMIAER-NIGSFSORAE-LGØGEETEFENULSOVHEVLPDIMSSEEFELKKINKTI 139 KORAPOLISILIKUSOPSATATEYMAALETAVGTYECOPDIAM-KERHTYODNGTU BAFL	212 209 204 205 90 90 90 90 274 248 253 70 265 265 265 265 265 269 269 269 271 268 268 263 263	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 hsPNNA5 hsPNNA10 mmPNNA10 hsPNNA11 GYPSY12_GAC mePNNAM51 mdPNNAM51 mdPNNAM51 hsPNNA1 hsPNNA2 mmPNNA3 hsPNNA4 mmPNNA3 mmPNNA3 hsPNNA4 mmPNNA3	340 -AL -LQAT 349 -Q -AL -LQAT 349 -Q -AL -CLQAT 349 -Q -AL -AL 349 -Q -AL -AL 349 -Q -AL -AL 349	357 3450 409 412 343 344 343 444 343 4443 445 112 353 353 364 453 353 365 351 450 453 360 453 390
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 hsPNMA10 hsPNMA10 hsPNMA11 GYPSY12_GAG mePNMA411 dPNMAM51 mdPNMA4 hsPNMA1 hsPNMA2 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNM3 hsPNM3 hsPNM3 hsPNM3 hsPNM3 hsPNM3 h	133 OPLL - EVMLYA-ELKVPSONT 1-510GALAPOANLEHT I BALQMOVPGOESKAKKINBEL 133 OPLL - EVMLYA-ELKVPSONT 1-510GALAPOANLEHT I BALQMOVPGOESKAKKINBEL 148 ESEKOENIYA-ELKVPSOTA 5-PSPGEETEFDULEOVTEIMPINVSEVEVRRELLESL 148 ESEKOENISTEVK-LUKVPSOTA 5-PSPGEETEFDULEOVTEIMPINVSEEVEVRRELLESL 135 - PIM-ASMAGP-NMKLPSORVV-PAOGEETEFNULTOVNGVLPDINMSEEFELKKINKTI 135 - PLV-ASMAGP-NMKLPSORVV-PAOGEETEFNULTOVNGVLPDINMSEEFELKKINKTI 135 - PLV-ASMAGP-NMKLPSORVV-PAOGEETEFNULSOVHEVLPDIMSSEEFELKKINKTI 135 - PLV-VKMAER-NMKLPSORVV-PAOGEETEFNULSOVHEVLPDIMSSEEFELKKINKTI 136 - PLI-AMIAER-MIQSFSORAE-LGØGEETEFNULSOVHEVLPDIMSSEEFELKKINKTI 137 - PLI-AMIAER-NIGSFSORAE-LGØGEETEFENULSOVHEVLPDIMSSEEFELKKINKTI 138 - PLI-AMIAER-NIGSFSORAE-LGØGEETEFENULSOVHEVLPDIMSSEEFELKKINKTI 139 KORAPOLISILIKUSOPSATATEYMAALETAVGTYECOPDIAM-KERHTYODNGTU BAFL	212 209 204 205 90 90 90 90 274 248 253 70 265 269 269 269 269 269 269 269 269 269 263 264 268 263 264 449 149	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 hsPNNA5 hsPNNA10 mmPNNA10 hsPNNA11 GYPSY12_GAG mePNNAM51 mdPNNAM51 hsPNNA2 hsPNNA1 hsPNNA2 mmPNNA4 hsPNNA4 mmPNNA3 mmP	340	357 345 350 409 412 551 342 343 343 343 343 464 443 443 353 352 365 351 352 365 351 352 365 351 352 365 353 365 351 352 390 390 390
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 mmPNMA10 hsPNMA11 GYPSY12_GAG mePNMA411 dPNMAM51 mdPNMA451 mdPNMA4 hsPNMA2 hsPNMA2 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA5 hsPNM5 hsPNM5 hsPNM5 hsPNM5 hsPNM5 hsPNM5 hsPNM5 hsPNM5 hsPNM5 hsPNM5 hsPNM5 hsPNM5 hsPNM5 hsPNM5 hsPNM5 hsPN5	133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEGENKKINKEL 133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPECKYKKINKEL 134 EPPK - ESMNYA - KLKVPSOT J- SPGETFEDNLEQVTEINFINQVSEVENRRLLES 148 KSILCMNSTCYK - KLKVPSOT J- SPGETFEDNLEQVTEINFINQVSEVENRRLLES 153 - PIM- ASMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NSMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NSMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NKMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 154 - PLI - ANTARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 155 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 156 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 157 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 158 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 159 - ROPAD LI RIIKL OTPDATVLEE O KLIDKQYGSTE - TPOQAYVQEL ITFOREG UPSEFV 160 - KGRAPD LI RIIKL VIDPATVLEE O KLIDKQYGSTE - TPOQAYVQEL ITFOREG UPSEFV 174 KGRADVIRILIK NNR DAT ITAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 276 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRI LIXANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORTAHOKFKI TURQOG VENG SFV 271 - ROPAREVIRU LAANAITTYKE CALAUQVFGYFS - SOXTAHOKFKI TURQOG VENG SFV 272 - ROPARE	212 209 204 205 90 90 90 90 90 274 248 253 70 265 269 269 269 269 269 269 269 269 269 269	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 mmPNNA5 hsPNNA10 mmPNNA10 hsPNNA11 GYPSY12_GAG mePNNAM51 mdPNNAM51 mdPNNAM51 hsPNNA1 msPNNA1 hsPNNA2 mmPNNA3 hsPNNA3 hsPNNA3 hsPNNA3 hsPNNA3 hsPNNA5 hsPNNA5 hsPNNA5	340 -AL -LQAT 349	357 3450 409 412 343 344 343 444 343 4443 445 112 353 353 364 453 353 365 351 450 453 360 453 390
mmPNMA3 hsPNMA5 mmPNMA5 mmPNMA10 mmPNMA10 hsPNMA11 mmPNMA11 GYPSY12_GAG mePNMAM51 mdPNMA451 mdPNMA451 mdPNMA451 mdPNMA2 mmPNMA2 mmPNMA2 mmPNMA3 mmPNMA3 mmPNMA3 mmPNMA4 mmPNMA4 mmPNMA4 mmPNMA5 mmPNMA5 mmPNMA5 mmPNMA5 mmPNMA10 mmPNMA10 mmPNMA11 sPSY12_GAG	133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEGENKKINKEL 133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPECKYKKINKEL 134 EPPK - ESMNYA - KLKVPSOT J- SPGETFEDNLEQVTEINFINQVSEVENRRLLES 148 KSILCMNSTCYK - KLKVPSOT J- SPGETFEDNLEQVTEINFINQVSEVENRRLLES 153 - PIM- ASMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NSMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NSMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NKMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 154 - PLI - ANTARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 155 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 156 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 157 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 158 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 159 - ROPAD LI RIIKL OTPDATVLEE O KLIDKQYGSTE - TPOQAYVQEL ITFOREG UPSEFV 160 - KGRAPD LI RIIKL VIDPATVLEE O KLIDKQYGSTE - TPOQAYVQEL ITFOREG UPSEFV 174 KGRADVIRILIK NNR DAT ITAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 276 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRI LIXANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORTAHOKFKI TURQOG VENG SFV 271 - ROPAREVIRU LAANAITTYKE CALAUQVFGYFS - SOXTAHOKFKI TURQOG VENG SFV 272 - ROPARE	212 209 204 205 90 90 90 274 245 253 70 269 269 269 269 269 269 269 261 264 149 149 149 149	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 hsPNNA5 hsPNNA10 mmPNNA10 hsPNNA11 GYPSY12_GAG mePNNAM51 mdPNNAM51 mdPNNAM51 mdPNNAM52 hsPNNA1 mmPNNA2 mmPNNA2 msPNNA4 mmPNNA3 msPNNA5 hsPNNA5 msPNN5 ms	340	357 345 350 409 412 551 342 343 343 343 343 464 443 443 353 352 365 351 352 365 351 352 365 351 352 365 353 365 351 352 390 390 390
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 mmPNMA10 hsPNMA11 GYPSY12_GAG mePNMA451 mdPNMA451 mdPNMA451 mdPNMA4 hsPNMA2 mmPNMA2 mmPNMA4 mmPNMA3 mmPNMA3 mmPNMA3 mmPNMA3 mmPNMA4 mmPNMA4 mmPNMA4 mmPNMA4 mmPNMA4 mmPNMA5 mmPNMA5 mmPNMA5 mmPNMA10 mmPNMA11 sPNMA411 sPNMA411 sPNMA411 sPNMA411	133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEGENKKINKEL 133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPECKYKKINKEL 134 EPPK - ESMNYA - KLKVPSOT J- SPGETFEDNLEQVTEINFINQVSEVENRRLLES 148 KSILCMNSTCYK - KLKVPSOT J- SPGETFEDNLEQVTEINFINQVSEVENRRLLES 153 - PIM- ASMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NSMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NSMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NKMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 154 - PLI - ANTARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 155 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 156 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 157 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 158 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 159 - ROPAD LI RIIKL OTPDATVLEE O KLIDKQYGSTE - TPOQAYVQEL ITFOREG UPSEFV 160 - KGRAPD LI RIIKL VIDPATVLEE O KLIDKQYGSTE - TPOQAYVQEL ITFOREG UPSEFV 174 KGRADVIRILIK NNR DAT ITAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 276 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRI LIXANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORTAHOKFKI TURQOG VENG SFV 271 - ROPAREVIRU LAANAITTYKE CALAUQVFGYFS - SOXTAHOKFKI TURQOG VENG SFV 272 - ROPARE	212 209 204 205 90 90 90 274 248 253 70 269 269 269 269 269 269 269 261 261 263 264 149 149 149 149	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 hsPNNA5 hsPNNA10 mmPNNA10 hsPNNA11 GYPSY12_GAG mePNNAM51 mdPNNAM51 mdPNNAM51 mdPNNA1 hsPNNA1 hsPNNA2 mmPNNA2 msPNNA4 mmPNNA3 hsPNNA5 msPNNA3 mmPNNA3 m	340	357 345 350 409 412 551 342 343 343 343 343 464 443 443 353 352 365 351 352 365 351 352 365 351 352 365 353 365 351 352 390 390 390
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 mmPNMA10 hsPNMA11 GYPSY12_GAG mePNMA411 dPNMAMS1 mdPNMA4 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA5 hsPNMA3 hsPNMA5 hsPNM5 hsPN5	133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEGENKKINKEL 133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPECKYKKINKEL 134 EPPK - ESMNYA - KLKVPSOT J- SPGETFEDNLEQVTEINFINQVSEVENRRLLES 148 KSILCMNSTCYK - KLKVPSOT J- SPGETFEDNLEQVTEINFINQVSEVENRRLLES 153 - PIM- ASMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NSMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NSMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NKMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 154 - PLI - ANTARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 155 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 156 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 157 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 158 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 159 - ROPAD LI RIIKL OTPDATVLEE O KLIDKQYGSTE - TPOQAYVQEL ITFOREG UPSEFV 160 - KGRAPD LI RIIKL VIDPATVLEE O KLIDKQYGSTE - TPOQAYVQEL ITFOREG UPSEFV 174 KGRADVIRILIK NNR DAT ITAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 276 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRI LIXANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORTAHOKFKI TURQOG VENG SFV 271 - ROPAREVIRU LAANAITTYKE CALAUQVFGYFS - SOXTAHOKFKI TURQOG VENG SFV 272 - ROPARE	212 209 209 209 205 90 90 274 248 253 70 265 265 269 269 269 269 269 269 261 268 268 264 268 264 264 264 263 264 263 264 249 249 264 265 269 269 269 269 269 269 209 209 209 209 209 209 209 209 209 20	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 hsPNNA5 hsPNNA10 mmPNNA10 hsPNNA11 GYPSY12_GAG mePNNAM51 mdPNNAM51 mdPNNA1 hsPNNA2 hsPNNA1 hsPNNA2 mmPNNA3 hsPNNA4 mmPNNA3 mmP	340	357 345 350 409 412 551 342 343 343 343 343 464 443 443 353 352 365 351 352 365 351 352 365 351 352 365 353 365 351 352 390 390 390
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 mmPNMA10 hsPNMA11 GYPSY12_GAG mePNMA411 dPNMAMS1 mdPNMAAS1 mdPNMAA hsPNMA3 hsPNMA1	133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEGENKKINKEL 133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPECKYKKINKEL 134 EPPK - ESMNYA - KLKVPSOT J- SPGETFEDNLEQVTEINFINQVSEVENRRLLES 148 KSILCMNSTCYK - KLKVPSOT J- SPGETFEDNLEQVTEINFINQVSEVENRRLLES 153 - PIM- ASMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NSMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NSMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NKMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 154 - PLI - ANTARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 155 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 156 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 157 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 158 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 159 - ROPAD LI RIIKL OTPDATVLEE O KLIDKQYGSTE - TPOQAYVQEL ITFOREG UPSEFV 160 - KGRAPD LI RIIKL VIDPATVLEE O KLIDKQYGSTE - TPOQAYVQEL ITFOREG UPSEFV 174 KGRADVIRILIK NNR DAT ITAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 276 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRI LIXANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORTAHOKFKI TURQOG VENG SFV 271 - ROPAREVIRU LAANAITTYKE CALAUQVFGYFS - SOXTAHOKFKI TURQOG VENG SFV 272 - ROPARE	212 209 209 209 209 205 90 90 274 248 253 70 265 265 265 265 265 269 269 269 269 269 264 264 264 264 264 264 264 149 149 149 149 306 306 306 306 318	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA10 mmPNNA5 hsPNNA10 mmPNNA10 hsPNNA11 GYPSY12_GAG mePNNAM51 mdPNNAM51 mdPNNA1 mmPNNA1 hsPNNA2 mmPNNA3 hsPNNA3 mmPNNA3 hsPNNA3 mmPNNA5 hsPNNA10 mmPNNA5 hsPNNA11 GYPSY12_GAG mePNNAM51 mdPNNAM51 mmPNNA1	340	357 345 350 409 412 551 342 343 343 343 343 464 443 443 353 352 365 351 352 365 351 352 365 351 352 365 353 365 351 352 390 390 390
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 mmPNMA10 hsPNMA11 GYPSY12_GAG mePNMA411 mdPNMA411 mdPNMA451 mdPNMA4 hsPNMA1 mmPNMA2 mmPNMA4 hsPNMA3 hsPNMA5 hsPNMA3 hsPNMA5 hsPNMA10 hsPNMA4 mmPNMA411 GYPSY12_GAG mePNMA451 mdPNMA451 mdPNMA451 hsPNMA11 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1	133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEGENKKINKEL 133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPECKYKKINKEL 134 EPPK - ESMNYA - KLKVPSOT J- SPGETFEDNLEQVTEINFINQVSEVENRRLLES 148 KSILCMNSTCYK - KLKVPSOT J- SPGETFEDNLEQVTEINFINQVSEVENRRLLES 153 - PIM- ASMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NSMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NSMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NKMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 154 - PLI - ANTARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 155 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 156 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 157 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 158 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 159 - ROPAD LI RIIKL OTPDATVLEE O KLIDKQYGSTE - TPOQAYVQEL ITFOREG UPSEFV 160 - KGRAPD LI RIIKL VIDPATVLEE O KLIDKQYGSTE - TPOQAYVQEL ITFOREG UPSEFV 174 KGRADVIRILIK NNR DAT ITAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 276 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRI LIXANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORTAHOKFKI TURQOG VENG SFV 271 - ROPAREVIRU LAANAITTYKE CALAUQVFGYFS - SOXTAHOKFKI TURQOG VENG SFV 272 - ROPARE	212 209 209 209 209 209 209 209 209 90 90 90 90 274 248 253 70 265 265 265 269 269 269 269 269 269 269 269 264 149 149 149 149 149 326 301 306 301 306 301 306 321 321	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 mmPNNA5 hsPNNA10 mmPNNA20 hsPNNA11 GYPSY12_GAG mdPNNAN51 mdPNNA41 mmPNNA1 hsPNNA2 mmPNNA3 mmPNNA3 hsPNNA3 mmPNNA3 mmPNNA3 hsPNNA5 hsPNNA1 mmPNNA3 mmPNNA3 hsPNNA5 hsPNNA1 mmPNNA3 hsPNNA5 hsPNNA1 mmPNNA1 hsPNNA1 hsPNNA1 hsPNNA1 hsPNNA1 hsPNNA1 hsPNNA1 hsPNNA1 hsPNNA1 hsPNNA1 hsPNNA3	340	357 345 350 409 412 551 342 343 343 343 343 464 443 443 353 352 365 351 352 365 351 352 365 351 352 365 353 365 351 352 390 390 390
mmPNMA3 hsPNMA5 mmPNMA10 mmPNMA10 hsPNMA11 GYP5Y12_GAG mePNMA411 mdPNMA451 mdPNMA451 mdPNMA451 hsPNMA3 mmPNMA2 mmPNMA2 mmPNMA3	133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEGENKKINKEL 133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPECKYKKINKEL 134 EPPK - ESMNYA - KLKVPSOT J- SPGETFEDNLEQVTEINFINQVSEVENRRLLES 148 KSILCMNSTCYK - KLKVPSOT J- SPGETFEDNLEQVTEINFINQVSEVENRRLLES 153 - PIM- ASMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NSMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NSMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 153 - PLV - NKMARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 154 - PLI - ANTARE - NMRL I SGREV - PAOGETFENNL I QVTGVL DDINMSEEE UKR INTI 155 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 156 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 157 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 158 - PLI - NILAER - NILSSGRAF - LOGENETFENNL I QVTGVL DDINMSEEE UKR INTI 159 - ROPAD LI RIIKL OTPDATVLEE O KLIDKQYGSTE - TPOQAYVQEL ITFOREG UPSEFV 160 - KGRAPD LI RIIKL VIDPATVLEE O KLIDKQYGSTE - TPOQAYVQEL ITFOREG UPSEFV 174 KGRADVIRILIK NNR DAT ITAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 276 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASSY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRIL KSTRATI TYAEC KALEQVFGSVE - SORDQI VELNTYDNEG USASYY 271 - ROPADVIRI LIXANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORDQI VELNTYDNEG USASYYY 271 - ROPAREVIRU LAANAITTYKE CALEQVFGSVE - SORTAHOKFKI TURQOG VENG SFV 271 - ROPAREVIRU LAANAITTYKE CALAUQVFGYFS - SOXTAHOKFKI TURQOG VENG SFV 272 - ROPARE	212 209 209 204 205 90 90 90 274 242 253 70 265 265 265 265 269 271 269 269 271 265 265 269 271 269 271 269 271 269 269 271 274 269 269 274 269 209 90 90 90 90 90 90 90 90 90 90 90 90 9	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 mmPNNA5 hsPNNA10 mmPNNA5 hsPNNA11 mmPNNA11 GYPSY12_GAG mdPNNAM51 mdPNNA451 mdPNNA45 hsPNNA1 mmPNNA2 mmPNNA3 hsPNNA3 mmPNNA5 hsPNNA4 mmPNNA5 hsPNNA4 mmPNNA5 hsPNNA11 mmPNNA5 hsPNNA4 mmPNNA5 hsPNNA5 mmPNNA5 hsPNNA4 mmPNNA5 hsPNNA4 mmPNNA5 hsPNNA1 mmPNNA5 hsPNNA5 hsPNNA1 mmPNNA5 hsPNNA5 mmPNNA5 hsPNNA4 mmPNNA5 hsPNNA4 mmPNNA5 hsPNNA1 hsPNNA5 hsPNNA1 hsPNNA5 mmPNNA5 hsPNNA3 mmPNNA5 hsPNNA4 mmPNNA5 hsPNNA4 mmPNNA5 hsPNNA5 hsPNNA5 mmPNNA5 hsPNNA5 mmPNNA5 hsPNNA5 mmPNNA5 hsPNNA5 mmPNNA5 hsPNNA5 mmPNNA5 hsPNNA4 mmPNNA5 hsPNNA5 hsPNNA5 mmPNNA5 hsPNNA5 mmPNNA5 hsPNNA5 hsPNNA5 hsPNNA5 hsPNNA5 hsPNNA5 hsPNNA5 mmPNNA5 hsPNN5 hsPNN5	340	357 345 350 409 412 551 342 343 343 343 343 464 443 443 353 352 365 351 352 365 351 352 365 351 352 365 353 365 351 352 390 390 390
mmPNMA3 mmPNMA5 mmPNMA10 mmPNMA10 hsPNMA11 GYP5Y12_GAG mePNMA411 mdPNMA451 mdPNMA451 mdPNMA451 mdPNMA4 hsPNMA2 mmPNMA2 mmPNMA2 mmPNMA3 mmPNMA3 mmPNMA3 mmPNMA5 mmPNMA5 mmPNMA5 mmPNMA5 mmPNMA5 mmPNMA10 mmPNMA10 mmPNMA11 GYPSY12_GAG mePNMA451 mdPNMA51 mdPNMA51 mdPNMA51 mdPNMA51 mdPNMA51 mdPNMA51 mdPNMA51 mdPNMA51 mdPNMA51 mdPNMA51 mdPNMA51 mdPNMA51 mdPNMA51 mdPNMA51 mdPNMA51 mdPNMA51 mdPNMA51 mdPNM451 mdP	133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEOSAWKKI MELL 133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEOSAWKKI MELL 134 EPPK - ESMNYA - KLKVPSOT J- SPGETFEDWL EQVT EINFINQVSEVU RRRLLES 135 - PIM - ASMADR - MIKL FSGRV - PAOGETFEDWL EQVT EINFINQVSEVU RRRLLES 135 - PIM - ASMADR - MIKL FSGRV - PAOGETFEDWL EQVT EINFINQVSEVU RRRLLES 135 - PLV - NSABRA - MIKL FSGRV - PAOGETFEDWL EQVT EINFINQSEEU KR.L MATL 135 - PLV - NSABRA - MIKL FSGRV - PAOGETFEDWL EQVT EINFINGVSEVU RRRLLES 136 - PLV - NSABRA - MIKL FSGRV - PAOGETFEDWL EQVT VL DDIWPSGEEU KR.L MATL 137 - PLI - ASMADR - MIKL FSGRV - PAOGETFEDWL EQVT VL DDIWPSGEEU KR.L MATL 138 - PLU - NKABRA - MIKL FSGRV - PAOGETFEDWL EQVT VL DDIWPSGEEU KR.L MATL 139 - RUFADLI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 198 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 198 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 198 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 108 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - STROD YN HELKTPOHK KR SKCV 207 RGPADVIRILK STRATT TAREK KALEDVFGSVE - SKDD YN HELTYDDING E USASYV 208 RGPADVIRILK STRATT TAREK KALEDVFGSVE - SKDD YN HELTYDDING E USASYV 211 RGPALD UNHIVQADNS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TTYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TTYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TYDROE E USASYV 211 RGPAREVN	212 209 209 209 209 209 90 90 90 274 248 253 70 265 269 269 269 269 269 269 269 269 269 269	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 hsPNNA5 hsPNNA10 hsPNNA10 hsPNNA11 GYPSY12_GAG mePNNAM51 mdPNNAM51 mdPNNAM51 msPNNA1 hsPNNA1 hsPNNA3 hsPNNA3 hsPNNA3 mmPNNA3 m	340	357 345 350 409 412 551 342 343 343 343 343 464 443 443 353 352 365 351 352 365 351 352 365 351 352 365 351 352 360 390 390
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 hsPNMA10 hsPNMA11 GYPSY12_GAG mePNMA411 sPNMA1 mdPNMAMS1 mdPNMAAS1 mdPNMAA hsPNMA2 mmPNMA1 hsPNMA3 hsPNMA5 hsPNMA5 mmPNMA5 hsPNMA5 mdPNMA51 mdPNMA4 hsPNMA5 hsPNMA1 mmPNMA1 mmPNMA1 mmPNMA1 mmPNMA1 hsPNMA3 mmPNMA3 hsPNMA5 hsPNMA1 mmPNMA1 hsPNMA3 hsPNMA1 mmPNMA1 hsPNMA1 mmPNMA1 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3	133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEOSAWKKI MELL 133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEOSAWKKI MELL 134 EPPK - ESMNYA - KLKVPSOT J- SPGETFEDWL EQVT EINFINQVSEVU RRRLLES 135 - PIM - ASMADR - MIKL FSGRV - PAOGETFEDWL EQVT EINFINQVSEVU RRRLLES 135 - PIM - ASMADR - MIKL FSGRV - PAOGETFEDWL EQVT EINFINQVSEVU RRRLLES 135 - PLV - NSABRA - MIKL FSGRV - PAOGETFEDWL EQVT EINFINQSEEU KR.L MATL 135 - PLV - NSABRA - MIKL FSGRV - PAOGETFEDWL EQVT EINFINGVSEVU RRRLLES 136 - PLV - NSABRA - MIKL FSGRV - PAOGETFEDWL EQVT VL DDIWPSGEEU KR.L MATL 137 - PLI - ASMADR - MIKL FSGRV - PAOGETFEDWL EQVT VL DDIWPSGEEU KR.L MATL 138 - PLU - NKABRA - MIKL FSGRV - PAOGETFEDWL EQVT VL DDIWPSGEEU KR.L MATL 139 - RUFADLI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 198 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 198 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 198 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 108 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - STROD YN HELKTPOHK KR SKCV 207 RGPADVIRILK STRATT TAREK KALEDVFGSVE - SKDD YN HELTYDDING E USASYV 208 RGPADVIRILK STRATT TAREK KALEDVFGSVE - SKDD YN HELTYDDING E USASYV 211 RGPALD UNHIVQADNS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TTYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TTYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TYDROE E USASYV 211 RGPAREVN	212 209 209 209 209 209 209 90 90 90 274 248 253 70 265 269 269 269 269 269 269 269 269 269 269	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 hsPNNA5 hsPNNA5 mmPNNA10 hsPNNA10 hsPNNA11 GYPSY12_GAG mePNNAM51 mdPNNAM51 mdPNNAM51 msPNNA1 hsPNNA2 hsPNNA3 mmPNNA3 m	340	357 345 350 409 412 551 342 343 343 343 343 464 443 443 353 352 365 351 352 365 351 352 365 351 352 365 351 352 360 390 390
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 mmPNMA10 hsPNMA11 GYPSY12_GAG mePNMA411 mdPNMA4S1 mdPNMA4S1 mdPNMA4 hsPNMA2 mmPNMA2 mmPNMA3 hsPNMA10 hsPNMA10 hsPNMA10 hsPNMA10 hsPNMA11 GYPSY12_GAG mePNMA451 mdPNMA451 mdPNMA451 hsPNMA1 hsPNMA1 hsPNMA3 mmPNMA3 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA3 mmPNMA3 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA3 mmPNMA3 hsPNMA4 hsPNMA3	133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEOSAWKKI MELL 133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEOSAWKKI MELL 134 EPPK - ESMNYA - KLKVPSOT J- SPGETFEDWL EQVT EINFINQVSEVU RRRLLES 135 - PIM - ASMADR - MIKL FSGRV - PAOGETFEDWL EQVT EINFINQVSEVU RRRLLES 135 - PIM - ASMADR - MIKL FSGRV - PAOGETFEDWL EQVT EINFINQVSEVU RRRLLES 135 - PLV - NSABRA - MIKL FSGRV - PAOGETFEDWL EQVT EINFINQSEEU KR.L MATL 135 - PLV - NSABRA - MIKL FSGRV - PAOGETFEDWL EQVT EINFINGVSEVU RRRLLES 136 - PLV - NSABRA - MIKL FSGRV - PAOGETFEDWL EQVT VL DDIWPSGEEU KR.L MATL 137 - PLI - ASMADR - MIKL FSGRV - PAOGETFEDWL EQVT VL DDIWPSGEEU KR.L MATL 138 - PLU - NKABRA - MIKL FSGRV - PAOGETFEDWL EQVT VL DDIWPSGEEU KR.L MATL 139 - RUFADLI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 198 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 198 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 198 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 108 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - STROD YN HELKTPOHK KR SKCV 207 RGPADVIRILK STRATT TAREK KALEDVFGSVE - SKDD YN HELTYDDING E USASYV 208 RGPADVIRILK STRATT TAREK KALEDVFGSVE - SKDD YN HELTYDDING E USASYV 211 RGPALD UNHIVQADNS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TTYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TTYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TYDROE E USASYV 211 RGPAREVN	212 209 209 209 209 209 209 209 209 90 90 90 274 248 253 265 265 265 269 269 269 269 269 269 269 269 269 269	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 hsPNNA5 hsPNNA5 mmPNNA10 hsPNNA10 hsPNNA11 GYPSY12_GAG mePNNAM51 mdPNNAM51 mdPNNAM51 msPNNA1 hsPNNA2 hsPNNA3 mmPNNA3 m	340	357 345 350 409 412 551 342 343 343 343 343 464 443 443 353 352 365 351 352 365 351 352 365 351 352 365 351 352 360 390 390
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 hsPNMA10 hsPNMA11 GYPSY12_GAG mePNMA411 mdPNMA451 mdPNMA451 mdPNMA4 hsPNMA2 mmPNMA2 hsPNMA3 mmPNMA5 hsPNMA10 hsPNMA11 GYPSY12_GAG mePNMA411 shSPNMA11 hsPNMA2 hsPNMA13 hsPNMA13 hsPNMA13 hsPNMA13 hsPNMA13 hsPNMA13 hsPNMA14 hsPNMA2 hsPNMA3 hsPNMA13 hsPNMA3	133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEOSAWKKI MELL 133 OPLL - EVML YA- ELKVPSONT J- SIGALAPOANLEH I ENLOWIQVPEOSAWKKI MELL 134 EPPK - ESMNYA - KLKVPSOT J- SPGETFEDWL EQVT EINFINQVSEVU RRRLLES 135 - PIM - ASMADR - MIKL FSGRV - PAOGETFEDWL EQVT EINFINQVSEVU RRRLLES 135 - PIM - ASMADR - MIKL FSGRV - PAOGETFEDWL EQVT EINFINQVSEVU RRRLLES 135 - PLV - NSABRA - MIKL FSGRV - PAOGETFEDWL EQVT EINFINQSEEU KR.L MATL 135 - PLV - NSABRA - MIKL FSGRV - PAOGETFEDWL EQVT EINFINGVSEVU RRRLLES 136 - PLV - NSABRA - MIKL FSGRV - PAOGETFEDWL EQVT VL DDIWPSGEEU KR.L MATL 137 - PLI - ASMADR - MIKL FSGRV - PAOGETFEDWL EQVT VL DDIWPSGEEU KR.L MATL 138 - PLU - NKABRA - MIKL FSGRV - PAOGETFEDWL EQVT VL DDIWPSGEEU KR.L MATL 139 - RUFADLI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 198 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 198 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 198 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - TPOQAYVOCH TTFOREG VD SFV 108 RGPAD LI RILK LOTPDATULEG OKLDKYNSTE - STROD YN HELKTPOHK KR SKCV 207 RGPADVIRILK STRATT TAREK KALEDVFGSVE - SKDD YN HELTYDDING E USASYV 208 RGPADVIRILK STRATT TAREK KALEDVFGSVE - SKDD YN HELTYDDING E USASYV 211 RGPALD UNHIVQADNS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPALD UNHIVQADNYS I SVECC EAFRQVFGSTE - SRRTS OVEYL RTYDREG E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TTYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TTYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TYDROE E USASYV 211 RGPAREVNEL UNANL TYNGEL ALUQVFGSTV - SNRTAU VYNL TYDROE E USASYV 211 RGPAREVN	212 209 209 209 209 209 209 209 90 90 90 274 248 265 265 265 269 269 271 269 269 269 271 265 269 269 271 265 269 269 271 265 269 269 271 265 269 269 271 265 265 269 271 265 265 269 271 265 265 265 269 271 265 269 271 274 245 265 265 269 271 269 271 274 245 265 265 269 271 274 245 265 265 269 271 265 269 271 265 269 271 274 265 269 271 265 269 271 269 269 271 269 269 271 269 269 271 269 269 271 269 269 271 269 269 271 269 269 271 269 269 271 269 269 271 269 269 271 269 269 271 269 269 271 269 269 271 269 269 269 271 269 269 269 271 265 269 269 271 265 269 269 271 265 269 269 271 265 265 265 269 269 271 226 326 269 271 226 326 263 263 264 149 149 149 221 306 221 326 269 226 226 269 227 269 226 269 226 269 226 269 226 269 226 269 226 269 226 269 226 269 226 269 226 269 226 269 226 269 226 269 226 269 226 269 226 31 260 226 31 226 326 269 226 31 226 326 269 226 31 226 326 322 269 226 31 221 312 20 312 20 312 20 20 20 20 20 20 20 20 20 20 20 20 20	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 mmPNNA5 hsPNNA10 mmPNNA10 hsPNNA11 GYPSY12_GAG mdPNNAM51 mdPNNA451 mdPNNA451 mdPNNA1 hsPNNA2 mmPNNA4 mmPNNA3 mmPNNA5 mdPNNA5 mdPNNA5 mdPNNA5 mdPNNA5 mmPNNA2 hsPNNA5 mmPNNA1 mmPNNA1 mmPNNA5 hsPNNA5 mmPNNA5 hsPNNA5 mmPNNA2 hsPNNA5 mmPNNA5 hsPNNA5 mmPNNA5 hsPNN5 hsPNA5 h	349	357 345 350 409 412 551 342 343 343 343 343 464 443 443 353 352 365 351 352 365 351 352 365 351 352 365 351 352 360 390 390
mmPNMA3 hsPNMA5 mmPNMA5 hsPNMA10 hsPNMA10 hsPNMA11 GYPSY12_GAG mePNMA411 mmPNMA1 mdPNMAMS1 mdPNMAAS1 mdPNMAAS1 mdPNMA3 hsPNMA3 hsPNMA3 hsPNMA3 hsPNMA5 hsPNMA5 hsPNMA5 mmPNMA1 mmPNMA1 mmPNMA1 hsPNMA5 hsPNMA1 mmPNMA1 hsPNMA3 hsPNMA1 mmPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA1 hsPNMA2 hsPNMA1 hsPNMA2 hsPNMA1 hsPNMA1 hsPNMA2 hsPNMA1 hsPNMA2 hsPNMA1 hsPNMA2 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3 hsPNMA3 mmPNMA3	133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQHMQVPGOSKAKKINAEL 133 OPLL - EVMLYA-ELKVPSONT 1-516ALAFUANLEHT I BALQHMQVPGOSKAKKINAEL 148 ESEKONTSTYK-KLKVPSOTA5-BSPGEFFETBULSQVTEIMPINVSEFVERREIHGEL 148 ESEKONTSTYK-KLKVPSOF5A-BSPGEFFETBULSQVTEIMPINVSEFVERREIHGEL 135 - PIM-ASMABA-NMKLFSGRVV-PAOGEFFETBULSQVTGVLPDHNMSEEFUKRIKNTI 135 - PIM-ASMABA-NMKLFSGRVV-PAOGEFFETBULSQVTGVLPDHNMSEEFUKRIKNTI 135 - PLV-VKMAER-NMKLFSGRVV-PAOGEFFETBULSQVTGVLPDHNMSEEFUKRIKNTI 135 - PLI-AMIAEN-NISSSGRAE-LGBGEFFENNISQVTEVLPDHSSEEFUKRIKNTI 216 RGPAADIVRFLKVSQPSATATEYMAALETAYGTTECGPDLMA-KFRHTYQDNGTUBAFL 216 RGPAADIVRFLKVSQPSATATEYMAALETAYGTTECGPDLMA-KFRHTYQDNGTUBAFL	212 209 209 209 209 209 209 90 90 90 274 248 253 70 265 269 269 269 269 269 269 269 269 269 269	mmPNNA4 hsPNNA3 mmPNNA3 hsPNNA5 hsPNNA5 hsPNNA10 mmPNNA10 hsPNNA11 GYPSY12_GAG mdPNNAM51 mdPNNAM51 hsPNNA31 mmPNNA2 mmPNNA3 mmPNNA3 mmPNNA1 hsPNNA5 hsPNNA10 mmPNNA1 hsPNNA5 hsPNNA10 mmPNNA1 hsPNNA1 hsPNNA5 mmPNNA1 hsPNNA1 hsPNNA1 hsPNNA1 hsPNNA1 hsPNNA1 hsPNNA1 hsPNNA3 mmPNNA3 hsPNNA5 hsPNNA3 hsPNA3 hsP	340	357 345 350 409 412 551 342 343 343 343 343 464 443 443 353 352 365 351 352 365 351 352 365 351 352 365 351 352 360 390 390

Figure 3. Multiple sequence alignment and phylogenetic tree of the *PNMA* family. (A) Multiple sequence alignment of the amino acid sequence of the Gag-like regions of marsupial *PNMA-MS1*, the human and mouse *PNMA* genes and Gypsy12_1_DR Gag. An evolutionarily conserved Gag-derived CX2CX4HX4C zinc finger motif is indicated by yellow shading. Residues conserved in all sequences are shaded black and highly conserved ones in grey. (B) A phylogenetic tree of *PNMA* family genes was constructed by the neighbour-joining method using the multiple alignment shown in Fig. 2. Bootstrap support (%) is shown for branches. *hs*: human:; *mm*: mouse; *md*: opossum; *me*: wallaby.

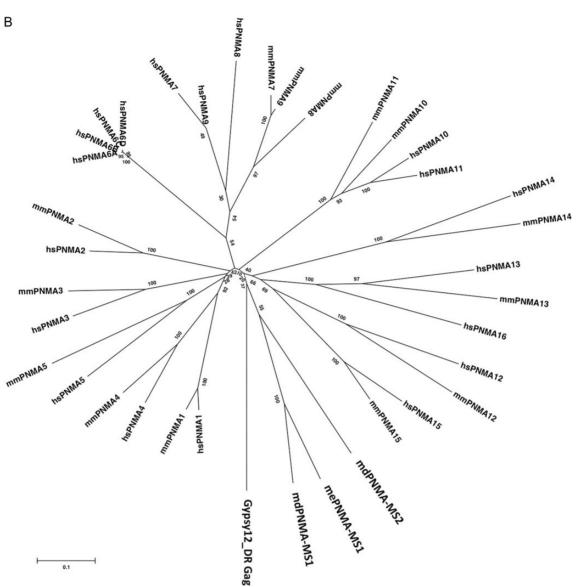


Figure 3. Continued

more closely related to the zebrafish Gypsy12_DR Gag than to the mouse and human proteins (Fig. 3B).

3.4. Comparative genomic analysis of PNMA-MS1 and -MS2

To elucidate whether *PNMA-MS1* is a marsupial-specific *PNMA* gene, comparative genomic analysis was performed using the VISTA tool with several vertebrate genomic sequences. *PNMA-MS1* was located in the intron 8 of the laminin alpha 3 (*LAMA3*) gene that is highly conserved in vertebrates. No *PNMA-MS1* orthologue was found in the syntenic region of any eutherian species, platypus (monotreme mammals), chicken (birds), frog (amphibian) and fugu (fish), demonstrating that *PNMA-MS1* is marsupial specific (Fig. 4A). It indicates that *PNMA-MS1* retrotransposition occurred only in the marsupial lineage after their divergence from eutherians (Fig. 5).

PNMA-MS2 was located between an *OR1Q1* gene and an *OR1J2*-like pseudogene (*ENSMODG00000 19710*) that lies 12-kb upstream of the former, in an OR gene cluster located between prostaglandinendoperoxide synthase 1 (*PTGS1*) and phosducin-like (*PDCL*) genes^{19,20} on opossum Chromosome 1. At present, we cannot confirm the presence or absence of *PNMA-MS2* in the tammar because the corresponding regions encompassing the *PTGS1* and *PDCL* genes are yet to be completely assembled. Recently, another Australian marsupial genome of the Tasmanian devil has been sequenced, and the region syntenic to that between opossum *OR1N2* and *PDCL* became available.

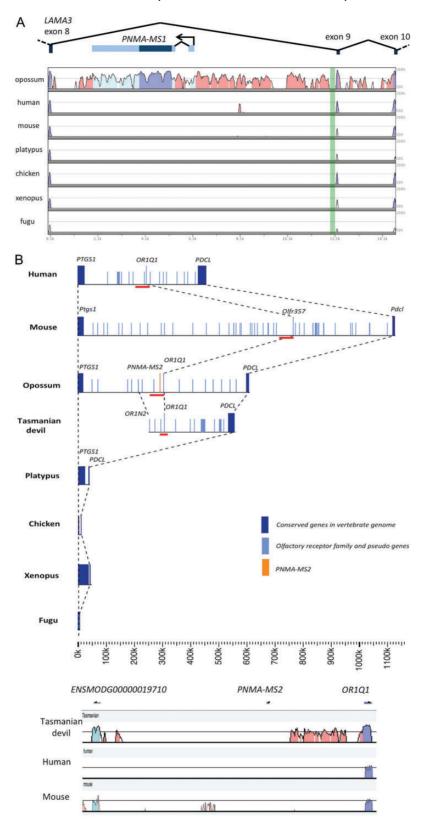


Figure 4. Comparative genomic analysis of the *PNMA-MS1* and *-MS2* regions in vertebrates. (A) *PNMA-MS1*. mLAGAN alignment of the tammar wallaby, opossum, human, mouse, platypus, chicken, frog and fugu *LAMA3* exons 8–10 region produced by mVISTA using the tammar sequence as the basis for comparison. Default parameters for mVISTA were used (conservation level, 70%, 100 bp window). Conserved regions appear as peaks highlighted in pink (>70% identity). Where these regions coincide with ORF sequences of *PNMA-MS1* or *LAMA3*, the peaks are shaded in purple. Where these regions coincide with the UTR region of *PNMA-MS1*, the peaks are shaded in light blue. The

It contains some gaps but none between opossum *OR1Q1* and *ENSMODG0000019710* (*OR1J2*-like pseudogene). A search of this region clearly demonstrated that the *PNMA-MS2* orthologue is absent from the Tasmanian devil genome.

The syntenic OR cluster lies between *PTGS1* and *PDCL* in the human Chromosome 9 and mouse Chromosome 2, respectively, but neither the number nor the order of OR genes and pseudogenes are conserved. As the human and mouse genome sequences in this region are complete and contain no gaps, the absence of the *PNMA-MS2* orthologue was confirmed (Fig. 4B). For the platypus, the *PTGS1* and *PDCL* genes are located next to each other, with no OR gene cluster and no *PNMA-MS2* orthologue, like in the chicken and fugu (Fig. 4B).

These results suggest that the integration of selected OR genes occurred between the *PTGS1* and *PDCL* genes in a common therian ancestor, and that the opossum-specific insertion of *PNMA-MS2* occurred after the divergence of eutherians and marsupials and the geographic separation of Australian and South American marsupials (Fig. 5). However, the possibility that the *PNMA-MS2* orthologue exists in some of Australian marsupial species cannot be excluded. It is possible that *PNMA-MS2* was deleted from Australian marsupial species after integration in a common marsupial ancestor (Fig. 5).

3.5. Expressions of PNMA-MS1 in the tammar wallaby and PNMA-MS2 in the opossum

PNMA-MS1 expression was investigated in several tissues in four different stages of the tammar wallaby, including foetal and pouch young stages. Human LAMA3 is expressed ubiquitously (EST profile Hs.436367). To exclude the possibility that heterogenous nuclear RNA (hnRNA) was detected between exons 8 and 9 of the LAMA3 gene rather than PNMA-MS1, we amplified *PNMA-MS1* using PCR primers designed within exons 1 and 2, respectively. Thus, the PCR product was shorter than its genomic sequence corresponding to hnRNAs of LAMA3 and PNMA-MS1. LAMA3 expression was analysed using primers designed to exons 79 and 81, near the 3'-UTR, due to the poor genome sequence quality in introns 84–90 of tammar LAMA3. Tammar LAMA3 expression was almost ubiquitous, with the exception of several pouch young tissues.

From Day 23 to 26 pregnancy, *PNMA-MS1* expression was detected in the foetal head and body, but there was

no expression in the YSP (Fig. 6A). In pouch young aged Day 60–70, PNMA-MS1 was detected only in the brain, kidney and ovary, but not in the liver, lung or testis (Fig. 6B). In Day 152 and 162 pouch young, changes in PNMA-MS1 expression were minimal, with expression detected in the kidney, liver, pancreas, heart, spleen and stomach, but not in the lung, bladder, adrenal, cerebrum or cerebellum (Fig. 6C). In the adult female, the brain (thalamus, hypothalamus and pituitary), ovary (ovary with active corpus luteum, corpus luteum alone, ovary with enlarged developing follicle and ovary with primary or secondary follicles), endometrium (gravid endometrium and non-gravid endometrium) and mammary gland (sucked gland and non-sucked gland) were examined. PNMA-MS1 expression was detected in the ovary (all four stages), mammary gland (sucked and non-sucked) and thalamus, but not in the hypothalamus or pituitary (Fig. 6D). There was no expression in either the gravid, or non-gravid, endometrium.

PNMA-MS2 expression was analysed in five tissues, the brain, liver, spleen, pituitary and ovary, using three primer sets designed in the putative coding frame. However, there was no expression in these tissues, suggesting that *PNMA-MS2* is not active in the opossum (data not shown). Although we cannot exclude the possibility that it may be expressed in a stage or tissue-specific manner, we conclude that *PNMA-MS2* is a pseudogene.

4. Discussion

4.1. PNMA-MS1 is a marsupial-specific PNMA gene

In this study, we have identified *PNMA-MS1* as a novel Ty3/Gypsy LTR retrotransposon-derived gene. Comparative genomic analysis showed that *PNMA-MS1* was present only in the marsupial lineage and was absent in the eutherian and monotreme mammals and in the non-mammalian vertebrates. *PEG10/SIRH1* and *SIRH12* are the only Ty3/Gypsy LTR retrotransposon-derived genes (derived from sushi-ichi-related retrotransposon-derived genes) reported in marsupials so far.^{11,21} Therefore, the *PNMA-MS1*, which is also a Ty3/gypsy LTR retrotransposon-derived gene, is the first and only member of the *PNMA* gene family in the marsupials.

The sushi-ichi-related retrotranposon that gave rise to the SIRH family was probably active around the

tammar genome sequence has a gap of about 200 bp that located at the 5'-side of exon 9 (shown by a green bar). (B) The region from *PTGS1* to *PDCL* in the opossum, human, mouse, platypus, chicken, frog and fugu genomes is shown (upper). *PNMA-MS2* is located between the *OR1Q1* gene and an *OR1J2*-like pseudgene (*ENSMODG00000019710*) in the opossum (red bar), but is absent in the Tasmanian devil, human and mouse genomes (red bars indicate the equivalent region in these species. An mVISTA alignment of the *OR1Q1*-*ENSMODG00000019710* region in the opossum (lower) confirms that *PNMA-MS2* is not present in the Tasmanian devil, human or mouse.

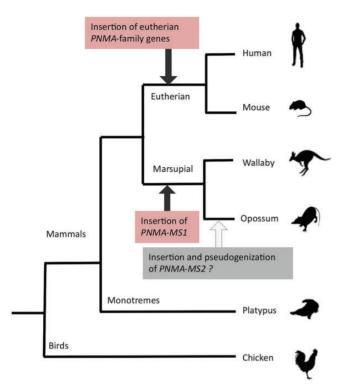


Figure 5. Evolutionary pathway of the PNMA family in mammals. PNMA-MS1 insertion occurred in a marsupial ancestor prior to the radiation of marsupial species. PNMA-MS2 was only found in the opossum and not in the Tasmanian devil, human and mouse, which suggests that PNMA-MS2 was acquired only in the opossum lineage after the divergence of the Australian and South American marsupials, but it does not show active transcription. The insertion of eutherian PNMA family genes occurred in a eutherian ancestor prior to the radiation of eutherian species. There are no PNMA family genes (or pseudogenes) in the platypus (monotremes) and chicken (birds).

time of the divergence between marsupials and eutherians, because PEG10/SIRH1 is conserved between the eutherians and marsupials,²¹ while SIRH3-11 seems to be eutherian-specific and SIRH12 evolved from a marsupial-specific retrotransposition event.¹¹ We did not detect any orthologues of PNMA-MS1 or -MS2 in the eutherian genome, nor any orthologues of eutherian PNMA1-16 genes in the marsupial genomes (data not shown). Due to many sequence gaps in the tammar wallaby, Tasmanian devil and opossum genomes, we cannot exclude the possibility that some marsupial orthologues of eutherian PNMA genes exist in such gap regions. Thus, it is possible that some retrotransposition events of Gypsy12_DR-related retrotranposon occurred in the common ancestor of the marsupials and eutherians. However, the higher similarity of PNMA-MS1 and -MS2 to Gypsy12_DR Gag than any other eutherian PNMAs suggests that their insertions in the marsupial genome were recent events. Taken together, these results suggest that the retrotransposition of Gypsy12_DR-related retrotransposon occurred after the divergence of the marsupials

and eutherians. The *PNMA* genes then evolved independently in these two lineages (Fig. 5).

In our analysis, only *PNMA-MS1* was detected in the tammar and opossum in contrast to 19 and 14 *PNMA* genes in humans and mice, respectively. We also observed the same trend in the *SIRH* genes: 11 genes in both humans and mice, while there are only 2 genes (*PEG10* and *SIRH12*) in the tammar genome.¹¹ This implies that the eutherian genome has a greater ability of exaptation as more Ty3/Gypsy types of LTR retrotransposons were incorporated into the genomes as endogenous genes than in the marsupial genomes.

4.2. The possible role of PNMA-MS1 genes in marsupial development

In rare cases, some retrotransposons have been incorporated as novel acquired genes into the host genomes and have contributed to the innovation of some eutherian-specific characteristics. Two such advantageous genes, *PEG10/SIRH1* and *PEG11/SIRH2*, play essential roles in the placental development in mice.^{7,8}

The role of *PNMA-MS1* in marsupial development and growth is less clear. *PNMA-MS1* expression was detected in the tammar brain, consistent with the expression of eutherian *PNMA* genes in brain. Interestingly, *PNMA-MS1* expression was confirmed only in the thalamus, but not in the hypothalamus, and pituitary in the adult brain. The thalamus has multiple functions including relaying sensation, spatial sense and motor signals to the cerebral cortex,²² so it is possible that *PNMA-MS1* is involved in the transmission of marsupial-specific sensations and signals. *PNMA-MS1* expression in the Day 60–70 pouch young ovary and adult female ovary suggests the gene may have a role in ovarian function. These issues will be addressed in a future study.

PNMA-MS1 protein has a conserved CCHC zinc finger domain. In retroviruses, this domain forms a part of the nucleocapsid protein that functions in virus genome packaging and the early infection process.²³ Proteins containing the CCHC zinc finger domain are commonly known to interact with single-stranded DNAs (ssDNAs) and RNAs.²⁴ The Drosophila Nanos protein is required for hunchback mRNA translational regulation in the early embryo to the establishment of the anterior-posterior body axis.²⁵ The mammalian cellular nucleic acid-binding protein (CNBP) containing seven CCHC domains is involved in neural crest development, affecting forebrain and craniofacial development.²⁶ CNBP has a single-stranded nucleic acid-binding ability and is also implicated in both transcriptional and translational regulations.^{27,28} Therefore, *PNMA-MS1* may also be involved in a specific-transcriptional or translational regulation by binding ssDNAs or RNAs.

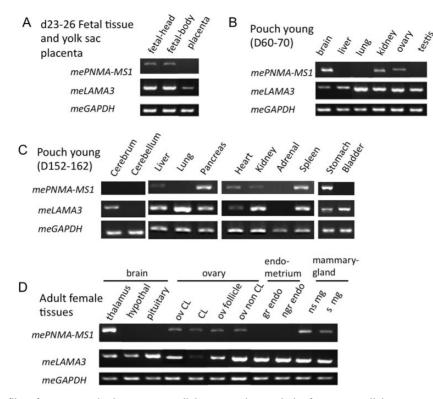


Figure 6. Expression profiles of *PNMA-MS1* in the tammar wallaby. Expression analysis of tammar wallaby *PNMA-MS1* in (A) foetal tissues and YSP at Day 23–26 of pregnancy, in several tissues, from (B) Day 60–70 and (C) Day 152–162 pouch young and (D) the adult female brain, ovary, endometrium and mammary gland. Expression of *GAPDH* and *LAMA3* for each sample is shown as a control. hypothal: hypothalamus; ov CL: ovary with active corpus luteum; CL: corpus luteum; ov follicle: ovary with developing follicle; ov non CL: ovary with primary or secondary follicle; gr endo: gravid endometrium; ngr endo: non-gravid endometrium; ns mg: non-sucked mammary gland; s mg: sucked (lactating) mammary gland.

5. Conclusions

We have identified one novel Ty3/Gypsy LTR retrotransposon-derived gene, PNMA-MS1 in marsupials as the first marsupial-specific PNMA gene reported. The high PNMA-MS1 expression levels in the thalamus, ovary and mammary gland provide intriguing questions as to its functions in marsupial development and growth as well as its role in marsupial evolution. Our data suggest that, in most of the cases, Ty3/Gypsy LTR retrotransposons have been independently incorporated into the marsupial and eutherian lineages. Consequently, the marsupials and eutherians have completely different sets of PNMA and SIRH genes, with the exception of PEG10. Thus, it is highly likely that these genes have evolved lineage-specific functions in the reproduction and development and contributed in establishing marsupial- or eutherian-specific traits, leading to the diversification of these two viviparous mammalian groups.

Acknowledgements: We thank Scott Brownlees, Bonnie Dopheide, Hongshi Yu, Jessica Stringer, Stephen Frankenberg and the Wallaby research group for the assistance with the animals and tissue collections, and Soichi Ogishima for his advice on phylogenetic analysis.

Supplementary Data: Supplementary Data are available at www.dnaresearch.oxfordjournals.org.

Accession Numbers

The National Center for Biotechnology Information GenBank (http://www.ncbi.nlm.nih.gov/Genbank) sequence accession number for tammar PNMA-MS1 is AB646689.

Funding

This work was supported by grants from the Australian Research Council Centre of Excellence in Kangaroo Genomics to M.B.R., Japan Society for the Promotion of Science (JSPS), Bilateral Program on Joint Research Project to F.I. and T.K-I. and Funding Program for Next Generation World-Leading Researchers (NEXT Program) to T.K-I., the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan, Grants-in-Aid for Scientific Research (S) and the Global Center for Excellence (GCOE) Program for International Research Center for Molecular Science in Tooth and Bone Diseases and Medical Research Institute, Tokyo Medical and Dental University, Joint Usage/Research Program to F.I.

References

- 1. Lander, E., Linton, L., Birren, B., et al. 2001, Initial sequencing and analysis of the human genome, *Nature*, **409**, 860–921.
- 2. Mouse Genome Sequencing ConsortiumWaterston, R.H., Lindblad-Toh, K., Birney, E., et al. 2002, Initial sequencing and comparative analysis of the mouse genome, *Nature*, **420**, 520–62.
- 3. Warren, W., Hillier, L., Marshall Graves, J., et al. 2008, Genome analysis of the platypus reveals unique signatures of evolution, *Nature*, **453**, 175–83.
- 4. Mikkelsen, T., Wakefield, M., Aken, B., et al. 2007, Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences, *Nature*, **447**, 167–77.
- 5. Renfree, M., Papenfuss, A., Deakin, J., et al. 2011, Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development, *Genome Biol.*, **12**, R81.
- Miller, K., Lynch, C., Martin, J., Herniou, E. and Tristem, M. 1999, Identification of multiple Gypsy LTR-retrotransposon lineages in vertebrate genomes, *J. Mol. Evol.*, **49**, 358–66.
- Ono, R., Nakamura, K., Inoue, K., et al. 2006, Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality, *Nat. Genet.*, **38**, 101–6.
- Sekita, Y., Wagatsuma, H., Nakamura, K., et al. 2008, Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta, *Nat. Genet.*, 40, 243–8.
- 9. Brandt, J., Veith, A. and Volff, J. 2005, A family of neofunctionalized Ty3/gypsy retrotransposon genes in mammalian genomes, *Cytogenet. Genome Res.*, **110**, 307–17.
- Youngson, N., Kocialkowski, S., Peel, N. and Ferguson-Smith, A. 2005, A small family of sushi-class retrotransposon-derived genes in mammals and their relation to genomic imprinting, *J. Mol. Evol.*, 61, 481–90.
- 11. Ono, R., Kuroki, Y., Naruse, M., et al. 2011, Identification of tammar wallaby SIRH12, derived from a marsupial-specific retrotransposition event, *DNA Res.*, **18**, 211–9.
- 12. Kaneko-Ishino, T. and Ishino, F. 2012, The role of genes domesticated from LTR retrotransposons and retroviruses in mammals, *Front Microbiol.*, **3**, 262.
- 13. Campillos, M., Doerks, T., Shah, P. and Bork, P. 2006, Computational characterization of multiple Gag-like human proteins, *Trends Genet.*, **22**, 585–9.
- 14. Takaji, M., Komatsu, Y., Watakabe, A., Hashikawa, T. and Yamamori, T. 2009, Paraneoplastic antigen-like 5 gene

(PNMA5) is preferentially expressed in the association areas in a primate specific manner, *Cereb. Cortex*, **19**, 2865–79.

- Rosenfeld, M., Eichen, J., Wade, D., Posner, J. and Dalmau, J. 2001, Molecular and clinical diversity in paraneoplastic immunity to Ma proteins, *Ann. Neurol.*, **50**, 339–48.
- Schüller, M., Jenne, D. and Voltz, R. 2005, The human PNMA family: novel neuronal proteins implicated in paraneoplastic neurological disease, *J. Neuroimmunol.*, **169**, 172–6.
- Poole, W.E., Simms, N.G., Wood, J.T. and Lubulwa, M. 1991, Tables for Age Determination of the Kangaroo Island Wallaby (Tammar), Macropus eugenii, from Body Measurements. Technical Note. Division of Wildlife and Ecology, CSIRO: Canberra, Australia.
- 18. NHMRC 2004, *Australian Code of Practice for the Care and Use of Animals for Scientific Purposes.* 7th edition. Australian Government: Canberra.
- 19. Godfrey, P., Malnic, B. and Buck, L. 2004, The mouse olfactory receptor gene family, *Proc. Natl Acad. Sci. USA*, **101**, 2156–61.
- 20. Malnic, B., Godfrey, P. and Buck, L. 2004, The human olfactory receptor gene family, *Proc. Natl Acad. Sci. USA*, **101**, 2584–9.
- 21. Suzuki, S., Ono, R., Narita, T., et al. 2007, Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting, *PLoS Genet.*, **3**, e55.
- 22. Ashwell, K. 2010, The neurobiology of Australian marsupials: brain evolution in the other mammalian radiation, Cambridge University Press, pp. 95–6.
- Narayanan, N., Gorelick, R. and DeStefano, J. 2006, Structure/function mapping of amino acids in the N-terminal zinc finger of the human immunodeficiency virus type 1 nucleocapsid protein: residues responsible for nucleic acid helix destabilizing activity, *Biochemistry*, 45, 12617–28.
- 24. Matsui, T., Kodera, Y., Endoh, H., et al. 2007, RNA recognition mechanism of the minimal active domain of the human immunodeficiency virus type-2 nucleocapsid protein, *J. Biochem.*, **141**, 269–77.
- Curtis, D., Treiber, D., Tao, F., Zamore, P., Williamson, J. and Lehmann, R. 1997, A CCHC metal-binding domain in Nanos is essential for translational regulation, *EMBO J.*, 16, 834–43.
- Chen, W., Liang, Y., Deng, W., et al. 2003, The zinc-finger protein CNBP is required for forebrain formation in the mouse, *Development*, **130**, 1367–79.
- Rajavashisth, T., Taylor, A., Andalibi, A., Svenson, K. and Lusis, A. 1989, Identification of a zinc finger protein that binds to the sterol regulatory element, *Science*, 245, 640–3.
- 28. Schlatter, S. and Fussenegger, M. 2003, Novel CNBP- and La-based translation control systems for mammalian cells, *Biotechnol. Bioeng.*, **81**, 1–12.