

Supporting Information

Cornelis et al. 10.1073/pnas.1417000112

SI Methods

Materials. All specimens of opossum (*M. domestica*) females were available from a breeding group kept at the animal facility of the Institute of Systematic Zoology at the Humboldt Universität Berlin, Museum of Natural History. This was performed under registration of the responsible State Office of Health and Social Affairs Berlin (LaGeSo), Germany. Placental and uterine tissues were collected from pregnant females killed at several gestation times and either stored in liquid nitrogen or processed for ISH. Other tissues (uterus, embryo, muscle, heart liver, lung, kidney, spleen, bladder, intestine, and ovary) were collected from the same killed females and from a nonpregnant female. Total RNA was extracted from the frozen organs by using the RNeasy RNA isolation kit (Qiagen). Near-term tammar wallaby placental tissues were collected and processed by B. Menzies and M. Renfree (University of Melbourne). Genomic DNA from *M. domestica* was extracted from the frozen tissues of a pregnant female. Genomic DNA from *Macropus rufus*, *Dendrolagus matschiei*, and *Phascolarctos cinereus* were purified from blood samples collected by B. Mulot and R. Potier (ZooParc de Beauval and Beauval Nature, Saint Aignan, France) by using the DNA Blood Kit II (PaxGene). Genomic DNA from *Potorous tridactylus*, *Pseudocheirus peregrinus*, *Perameles gunnii*, *Isodon macrourus*, *Dasyurus geoffroyi*, *Sminthopsis crassicaudata*, and *Didelphis virginiana* were provided by M. Nilsson and A. Janke, LOEWE Biodiversity and Climate Research Center. Genomic DNA from *Caluromys philander*, *Hyladelphys kalinowskii*, *Marmosops parvidens*, *Metachirus nudicaudatus*, *Philander opossum*, *Didelphis marsupialis*, *M. murina*, *M. demerarae*, and *M. brevicaudata* were extracted from tissues provided by F. Catzeflis, Collection de Tissus de Mammifères de Montpellier, Montpellier, France. Genomic DNA from *M. emiliae*, *M. theresa*, *M. adusta*, and *M. americana* were extracted from tissues provided by C. Conroy, Museum of Vertebrate Zoology, University of California, Berkeley, CA. Unless indicated otherwise, genomic DNAs were extracted from tissues conserved in ethanol. Genomic DNAs were purified by phenol–chloroform extraction.

Real-Time RT-PCR. *Env* mRNA expression was determined by qRT-PCR. Reverse transcription was performed with 100 ng of DNase-treated RNA as in ref. 1. PCR was carried out with 5 μ L of diluted (1:20) cDNA in a final volume of 25 μ L by using the FastSYBR Green PCR Master Mix (Qiagen) in an ABI PRISM 7000 sequence detection system. Primers are listed in Table S2. Transcript levels were normalized relative to the amount of the housekeeping RPLP0 gene (ribosomal protein P0). Samples (three samples per organ) were assayed in duplicate. RACE-PCR was performed with 100 ng of DNase-treated RNA, by using the SMARTer RACE cDNA Amplification Kit (Clontech).

ISH. Freshly collected *M. domestica* placentae were fixed in 4% (wt/vol) paraformaldehyde at 4 °C, embedded in paraffin, and serial

sections (7 μ m) were either stained with hematoxylin-eosin-saffron (HES) or used for ISH. For *syncytin-Opo1* and *env3*, four PCR-amplified fragments of ~400 bp (primers listed in Table S2) were cloned into pGEM-T Easy (Promega) for in vitro synthesis of the antisense and sense riboprobes, generated with SP6 RNA polymerase and digoxigenin 11-UTP (Roche Applied Science) after cDNA template amplification. Sections were processed, hybridized at 42 °C overnight with the pooled riboprobes, and incubated further at room temperature for 2 h with alkaline phosphatase-conjugated antidigoxigenin antibody Fab fragments (Roche Applied Science). Staining was revealed with nitroblue tetrazolium and 5-bromo-4-chloro-3-indoyl phosphate phosphatase alkaline substrates, as indicated by the manufacturer (Roche Applied Science).

Electron Microscopy. Tissues were fixed at least 1 h at 4 °C in 2.5% (vol/vol) glutaraldehyde in 0.1 M Sørensen phosphate buffer (pH 7.3), postfixed 2 h in aqueous 2% (wt/vol) osmium tetroxide, stained en bloc in 2% (wt/vol) uranyl acetate in 30% (vol/vol) methanol, dehydrated, and finally embedded in Epon. Analysis of the cellular organization of the tissue at the optical level was performed with 1- μ m semithin sections. Mounted on a microscope slide, the sections were de-embedded for 20 min at room temperature in sodium hydroxide-saturated ethanol and stained for 30 min at 56 °C with a 1% aqueous solution of thionine blue. After washings in absolute ethanol and xylene, the sections were mounted with Eukitt mounting medium. For ultrastructural observation, 70-nm ultrathin sections were stained with uranyl acetate and lead citrate and examined with a FEI Tecnai 12 microscope at 80 Kv.

Syncytin-Opo1 Expression Vector and Fusion Assay. The *syncytin-Opo1* fragments PCR-amplified from the genomic DNA of each *Monodelphis* species possessing the *syncytin-Opo1* ortholog were digested with EcoRI and MluI, and the PCR products were cloned into the pHCMV-G vector (GenBank accession no. AJ318514; gift from F.-L. Cosset, Ecole Normale Supérieure, Lyon, France). The *M. domestica syncytin-Opo1* gene was also synthesized and codon-optimized for its expression in human cells (GenCust; GenBank accession no. KM235330). Cell–cell fusion assays were performed by cotransfecting 293T cells ($1\text{--}2 \times 10^5$ cells per well) with *syncytin-Opo1*-expressing vectors along with a vector expressing an nls-LacZ gene (0.5–1 μ g at a ratio of 1:1) using the Lipofectamine LTX and Plus reagent transfection kit (Invitrogen), before coculturing them (1:1 ratio) with a panel of target cells. At 24–48 h after transfection, cells were fixed and X-Gal-stained. Fusion index was calculated as $[(N - S)/T] \times 100$, where N is the number of nuclei in the syncytia, S is the number of syncytia, and T is the total number of nuclei counted. All cell lines are described in refs. 2 and 3, except for the opossum OK cell line (ATCC CRL-1840). They were grown in DMEM supplemented with 10% (vol/vol) FCS (Invitrogen), 100 mg/mL streptomycin, and 100 U/mL penicillin.

1. de Parseval N, Lazar V, Casella JF, Benit L, Heidmann T (2003) Survey of human genes of retroviral origin: Identification and transcriptome of the genes with coding capacity for complete envelope proteins. *J Virol* 77(19):10414–10422.

2. Dupressoir A, et al. (2005) Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. *Proc Natl Acad Sci USA* 102(3):725–730.

3. Cornelis G, et al. (2012) Ancestral capture of syncytin-Car1, a fusogenic endogenous retroviral envelope gene involved in placentation and conserved in Carnivora. *Proc Natl Acad Sci USA* 109(7):E432–E441.

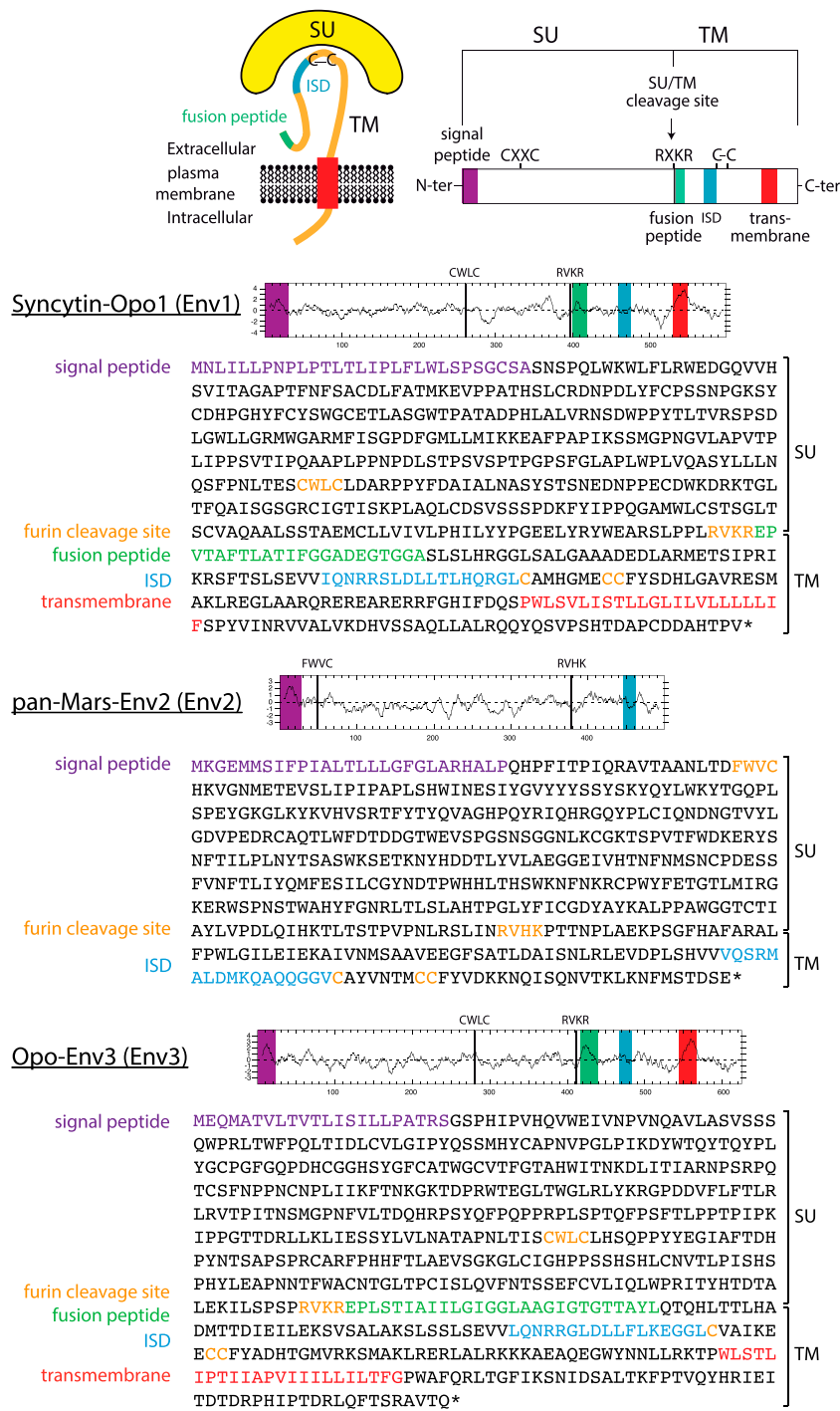


Fig. S1. Primary sequences of syncytin-Opo1, pan-Mars-Env2, and Opo-Env3. Primary amino acid sequences and characteristic structural features of the opossum Env proteins are provided, together with their hydrophobicity plots. The same color code and abbreviations as in Fig. 2 are used.

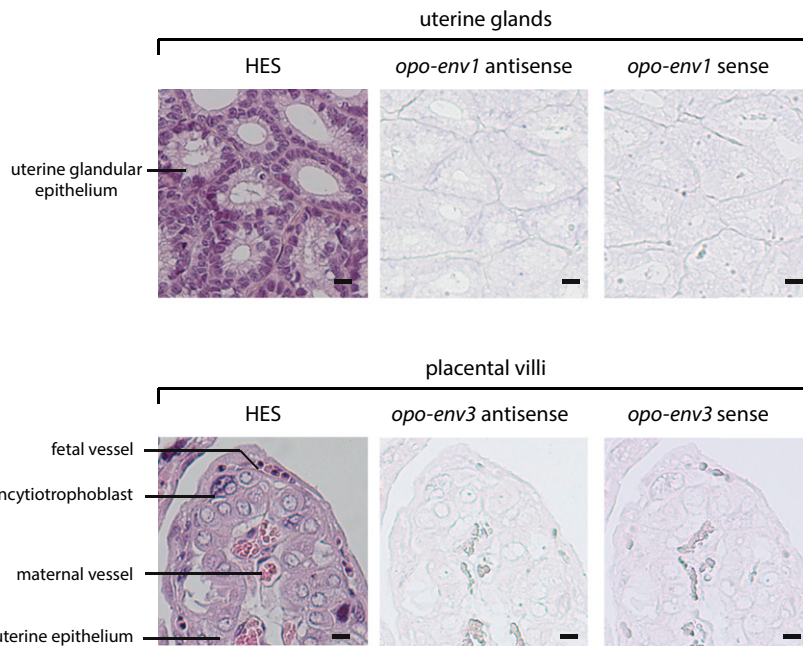


Fig. S2. Control ISH for lack of expression of *opo-env1* and *-env3* on uterine glands or placental villi sections, respectively. HES-stained sections of placenta and ISH on serial sections for *opo-env1* (Upper) or *opo-env3* (Lower) placental transcripts using digoxigenin-labeled antisense or sense riboprobes revealed with an alkaline phosphatase-conjugated antidigoxigenin antibody. Uterine glands (Upper) and placental villi (Lower) from the same hybridized sections as in Fig. 4. No staining is observed for either gene. (Scale bars: 10 μm .)

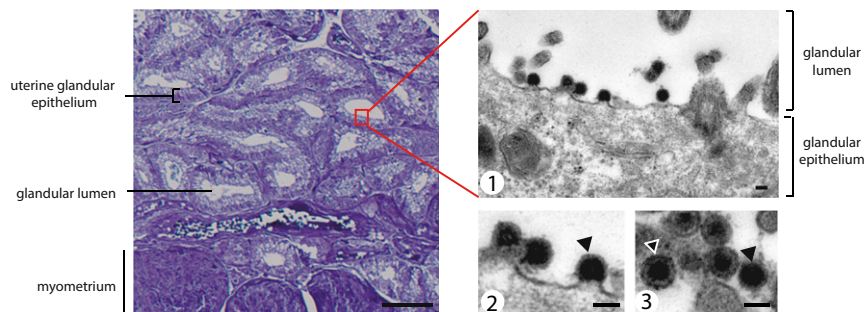


Fig. S3. Semithin sections and electron microscopy analysis of maternal uterine glands displaying retroviral-like particles budding in the glandular lumen. (Left) Semithin section of maternal uterine glands, with the uterine glandular epithelium, glandular lumen, and myometrium indicated. (Scale bars: 50 μm .) (Right) Electron microscopy of viral "immature" particles budding at the cell surface of a glandular epithelial cell into the glandular lumen (1 and 2), and released viral particles either immature (black arrowhead) or mature (black and white arrowheads) with an icosahedral condensed core (3). (Scale bar: 100 nm.)

Syncytin-Opo1

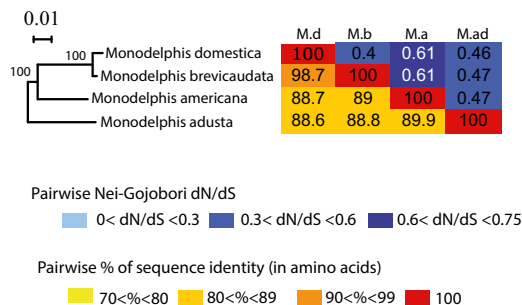


Fig. S4. Sequence conservation and evidence for purifying selection of syncytin-Opo1 in *Monodelphis* species. (Left) Syncytin-Opo1 phylogenetic tree determined by using amino acid alignment of the encoded proteins identified in Fig. 7, by the maximum-likelihood method. The horizontal branch length and scale indicate the percentage of amino acid substitutions. Percent bootstrap values obtained from 1,000 replicates are indicated at the nodes. (Right) Double-entry table for the pairwise percentage of amino acid sequence identity between the *syncytin-Opo1* gene among the indicated species (lower triangle) and the pairwise Nei-Gojobori nonsynonymous-to-synonymous mutation rate ratio (dN/dS; upper triangle). A color code is provided for both series of values.

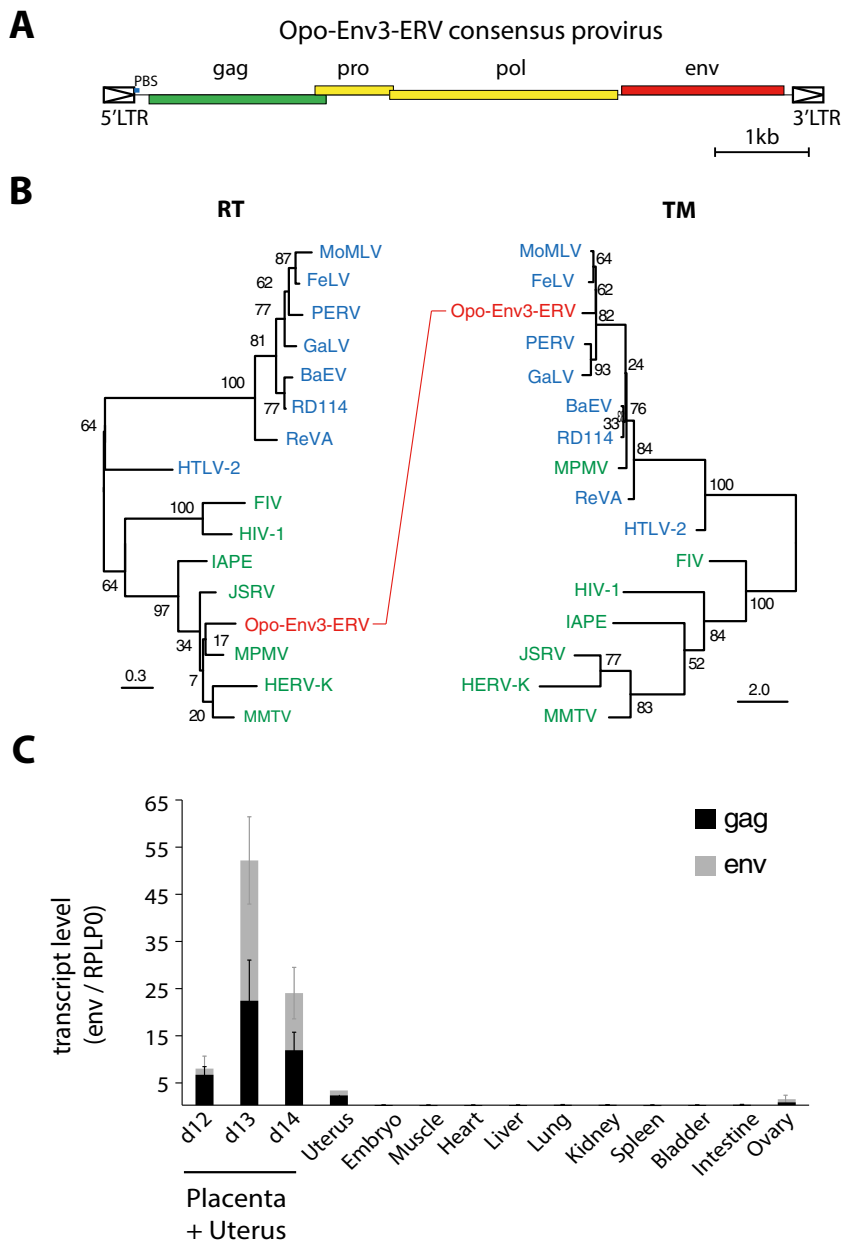


Fig. 56. Genetic structure and expression of the Opo-Env3-containing ERV. (A) Genetic structure of the consensus Opo-Env3-ERV provirus generated from the sequences of the 11 coding opo-env3 genes associated proviruses. This consensus provirus discloses full-length coding gag-pro-pol and env genes, as well as >92% identical LTRs and a PBS most closely related to the complementary sequence of the opossum Lys tRNA. (B) Opo-Env3-ERV is a mosaic retrovirus with a betaretrovirus *pol* gene and a gammaretrovirus *env* gene. The phylogenetic trees were determined as described in ref. 1 by using the catalytic domain sequences of the RT region of the indicated *pol* genes (Left) or the TM region of the corresponding *env* genes (Right) by the maximum-likelihood method, using the consensus sequence for the Opo-Env3-ERV genes. The horizontal branch length and scale indicate the percentage of amino acid substitutions. Percent bootstrap values obtained from 1,000 replicates are indicated at the nodes. (C) Real-time qRT-PCR analysis of the Opo-Env3-ERV *gag* (black) and *env* (gray) gene transcripts from the opossum (*M. domestica*). Transcript levels are expressed as the ratio of the expression level of each retroviral gene to that of the *RPLP0* control gene (Methods). Because of the high interpenetration of maternal and fetal tissues, placental and uterine tissues are analyzed as a whole at three gestational dates (day 12, 13, and 14). Values are the means of duplicates from three samples \pm SEM.

1. Bénit L, Dessen P, Heidmann T (2001) Identification, phylogeny, and evolution of retroviral elements based on their envelope genes. *J Virol* 75(23):11709–11719.

Table S1. *M. domestica* Env protein coding sequences genomic coordinates

Env coding gene	Chromosome	Strand orientation	Coordinates
Env1	chr8	-	120625237-120627027
Env2	chr5	-	11260853-11262349
Env3 (chr1)	chr1	+	242599227-242600672
Env3 (chr2-1)	chr2	-	52421962-52423827
Env3 (chr2-2)	chr2	-	259517711-259519216
Env3 (chr2-3)	chr2	-	269216373-269217734
Env3 (chr3)	chr3	+	526989485-526991362
Env3 (chr4)	chr4	-	412041177-412042661
Env3 (chr6-1)	chr6	+	90633297-90635162
Env3 (chr6-2)	chr6	+	229324539-229326416
Env3 (chr6-3)	chr6	+	258334506-258336437
Env3 (chrUn-1)	chrUn	+	41432161-41434035
Env3 (chrUn-2)	chrUn	+	49353246-49355111
Env4 (chr1)	chr1	+	378680256-378681854
Env4 (chr2-1)	chr2	-	203870817-203872415
Env4 (chr2-2)	chr2	+	251101334-251102830
Env4 (chr3-1)	chr3	-	368911730-368914405
Env4 (chr3-2)	chr3	+	392091104-392093905
Env4 (chr3-3)	chr3	+	392097607-392100264
Env4 (chr8)	chr8	-	169483977-169485383
Env5 (chr4-1)	chr4	+	85997868-85999661
Env5 (chr4-2)	chr4	+	433659095-433660816
Env6 (chr3)	chr3	-	55370589-55371965
Env6 (chr6)	chr6	-	80144835-80146850
Env7 (chr1)	chr1	-	148492139-148494058
Env7 (chr3)	chr3	+	115079115-115080992
Env7 (chr4-1)	chr4	-	331702780-331704207
Env7 (chr4-2)	chr4	+	433659095-433660816
Env8 (chr1-1)	chr1	-	14632900-14634915
Env8 (chr1-2)	chr1	-	22458558-22460573
Env8 (chr1-3)	chr1	-	44922788-44924803
Env8 (chr1-4)	chr1	-	110916197-110918212
Env8 (chr1-5)	chr1	-	128115494-128117509
Env8 (chr1-6)	chr1	+	143726215-143728230
Env8 (chr1-7)	chr1	-	167895124-167897139
Env8 (chr1-8)	chr1	+	195056807-195058591
Env8 (chr1-9)	chr1	-	327576689-327578680
Env8 (chr1-10)	chr1	-	408682997-408684655
Env8 (chr1-11)	chr1	+	414372207-414374222
Env8 (chr1-12)	chr1	-	460528976-460530991
Env8 (chr1-13)	chr1	+	468658869-468660884
Env8 (chr1-14)	chr1	+	604520089-604522080
Env8 (chr1-15)	chr1	+	663273757-663275772
Env8 (chr2-1)	chr2	-	134198224-134200239
Env8 (chr2-2)	chr2	+	196643545-196645560
Env8 (chr2-3)	chr2	-	260753341-260754750
Env8 (chr2-4)	chr2	-	261525352-261527367
Env8 (chr2-5)	chr2	+	262350119-262354627
Env8 (chr2-6)	chr2	-	423468966-423470981
Env8 (chr2-7)	chr2	-	472324337-472326073
Env8 (chr3-1)	chr3	-	539612-541627
Env8 (chr3-2)	chr3	-	11971855-11973870
Env8 (chr3-3)	chr3	-	20146270-20151123
Env8 (chr3-4)	chr3	-	29779101-29781116
Env8 (chr3-5)	chr3	+	334022522-334024537
Env8 (chr3-6)	chr3	-	439830531-439832540
Env8 (chr4-1)	chr4	-	28979732-28981747
Env8 (chr4-2)	chr4	+	111516459-111518474
Env8 (chr4-3)	chr4	+	335872638-335874653
Env8 (chr4-4)	chr4	+	407540208-407542223
Env8 (chr4-5)	chr4	+	414607752-414609581
Env8 (chr4-6)	chr4	-	416947942-416949957
Env8 (chr5-1)	chr5	+	38428969-38430984

Table S1. Cont.

Env coding gene	Chromosome	Strand orientation	Coordinates
Env8 (chr5-2)	chr5	–	103123482–103125497
Env8 (chr5-3)	chr5	–	153234093–153236108
Env8 (chr5-4)	chr5	+	213799757–213801772
Env8 (chr6-1)	chr6	+	17978209–17980224
Env8 (chr6-2)	chr6	+	34339376–34341391
Env8 (chr6-3)	chr6	–	79468511–79470526
Env8 (chr8)	chr8	+	155112001–155114016
Env8 (chr-Un1)	chrUn	+	64499316–64501331
Env8 (chr-Un2)	chrUn	–	65322701–65324716
Env8 (chr-Un3)	chrUn	+	82972007–82973659
Env8 (chr-Un4)	chrUn	+	99851075–99852727
Env8 (chrX-1)	chrX	+	25699343–25701358
Env8 (chrX-2)	chrX	–	25900813–25902828
Env9	chr3	–	504255693–504258332

Chromosome number and coordinates are given according to the reference Opossum genome (monDom5) from the University of California, Santa Cruz genome database (genome-euro.ucsc.edu/cgi-bin/hgGateway?hgid=198434456_mYJHBqra3j3ET9GlcNBaOQ28GO3b&clade=mammal&org=Opossum&db=0).

Table S2. List of primers

	Primers
qRT-PCR	
RPLP0-F	5'-AAG-ACA-GGG-CTA-CCT-GGA-A
RPLP0-R	5'-ATC-TGC-TGC-ATC-TGC-TTG-G
Opo-Env1-qRT-PCR-F	5'-GAC-TGG-CCT-CCT-TAC-ACT-TT
Opo-Env1-qRT-PCR-R	5'-CAT-AGA-GCT-TTT-GAT-TGG-G
Opo-Env2-qRT-PCR-F	5'-GCT-GAA-GGG-GGA-GAG-ATT
Opo-Env2-qRT-PCR-R	5'-GCC-AAG-GGG-TGT-CAT-TAT
Opo-Env3-qRT-PCR-F	5'-CAC-TCC-CAT-CTT-TGT-AAT-GTT-AC
Opo-Env3-qRT-PCR-R	5'-AGT-GTT-GCA-GGC-CCA-AAA
Opo-Env4-qRT-PCR-R	5'-CTC-CTC-CTG-CTG-GCT-GTG
Opo-Env4-qRT-PCR-R	5'-TCC-ATC-CAA-TTG-CAA-TTA-GA
Opo-Env5-qRT-PCR-F	5'-GAT-ATA-GCT-AGG-CTA-GAA-TCC-TC
Opo-Env5-qRT-PCR-R	5'-GGT-TAA-CAT-AGA-AGC-AGC-ATT
Opo-Env6-qRT-PCR-F	5'-TCT-GAC-CAC-CAG-CTG-CTG
Opo-Env6-qRT-PCR-R	5'-GCA-AGG-CAA-TTG-TTG-GGC
Opo-Env7-qRT-PCR-F	5'-CAA-AAC-CCC-AGA-TGG-AAA
Opo-Env7-qRT-PCR-R	5'-GGG-ATA-TGT-GTT-GGG-AAG-TTA-G
Opo-Env8-qRT-PCR-F	5'-CGA-GAC-CAG-GAC-GAG-AAG
Opo-Env8-qRT-PCR-R	5'-GAG-TAA-ATG-TAA-TGT-CAG-TGT-TC
Opo-Env9-qRT-PCR-F	5'-CTG-TGT-GCT-TTG-GAC-ATT-CTA
Opo-Env9-qRT-PCR-R	5'-GCT-CTA-TGA-AAG-TTC-TGT-TGG-TTA
Opo-Env3-ERV-Gag-qRT-PCR-F	5'-AAG-CCG-ACA-CCA-ATA-AAC-AGG
Opo-Env3-ERV-Gag-qRT-PCR-R	5'-GTG-GTA-TTT-CCT-CCT-TCT-CAA
5' RACE	
Opo-Env1-5'RACE-R	5'-GAT-TAG-GGT-TAG-GGT-TGG-AAG-AGG-GTT-G
Opo-Env1-3'RACE-F	5'-AGT-CCC-ACG-TCC-AAA-AAG-CAG-AA
Opo-Env3-5'RACE-Gag-R	5'-ATA-CAT-ACG-CGT-ATT-TTA-TCA-GCC-TGG-AGT-GC
Opo-Env3-5'RACE-Env-R	5'-TGA-TCT-GGT-GGC-TGG-TAG-CAG-GAT-AGA-G
ISH	
Opo-Env1-ISH-1-F	5'-GCC-CGC-ATG-TTT-ATC-TCC
Opo-Env1-ISH-1-R	5'-TGT-CCT-TCC-AGT-CAC-ACT-CG
Opo-Env1-ISH-2-F	5'-CCC-GCA-TCA-AAA-GGT-CCT
Opo-Env1-ISH-2-R	5'-AAC-GGA-CTG-GTA-TTG-TTG-CC
Opo-Env1-ISH-3-F	5'-ATC-CTT-CTT-CCC-AAC-CCT
Opo-Env1-ISH-3-R	5'-AAA-GTG-TAA-GGA-GGC-CAG-TC
Opo-Env1-ISH-4-F	5'-CTC-ACC-TTT-CAG-GCC-ATC
Opo-Env1-ISH-4-R	5'-GTC-TCC-ATC-CTG-GCC-AAG
Opo-Env3-ISH-1-F	5'-GGC-GAC-AGT-TCT-GAC-AGT-AAC
Opo-Env3-ISH-1-R	5'-GAT-TAA-AGG-AGC-AGG-TCT-GAG
Opo-Env3-ISH-2-F	5'-AAA-TTC-ACC-AAC-AAG-GG
Opo-Env3-ISH-2-R	5'-GCA-GAG-GTG-TTA-TAA-GGG-TGA
Opo-Env3-ISH-3-F	5'-CCC-ATC-TTT-GTA-ATG-TTA-C
Opo-Env3-ISH-3-R	5'-TAC-AAC-TTC-TGA-GAG-GGA-GGA
Opo-Env3-ISH-4-F	5'-GCC-TCG-ACC-TAC-TTT-TTT-TAA
Opo-Env3-ISH-4-R	5'-AGA-CGA-TCA-GTA-GGG-ATG-TGT
Amplification of Opo-Env in marsupials	
syncytin-Opo1-locus-F	5'-CRC-TGR-AAT-KTA-AAG-RGA-TCA-G
syncytin-Opo1-locus-R	5'-ATA-CAT-ACG-CGT-GCC-CTT-TAC-ATT-TTC-CAA-GAA-TA
syncytin-Opo1-ORF-F	5'-ATA-CAT-GAA-TTC-ATC-CAG-CAA-ACC-CCC-TGA
syncytin-Opo1-ORF-R	5'-ATA-CAT-ACG-CGT-GCA-ATC-TCC-GCT-TCT-TCT-AC
syncytin-Opo1-Int-F	5'-CTC-TGC-TTC-AAA-TTC-ACC-TC
syncytin-Opo1-Int-R	5'-ATG-CCT-CCT-TCT-TAA-TCA
Opo-Env3-ORF-F	5'-ATA-CAT-CTC-GAG-AAT-ATC-CCC-TCC-CCT-TAA-T
Opo-Env3-ORF-R	5'-ATA-CAT-ACG-CGT-AGG-CTA-CAC-AGG-GGG-AGT
Opo-Env3-locus-chr2-F	5'-GAC-CAC-AGG-ATG-ATA-AAG-TTT
Opo-Env3-locus-chr2-R	5'-CTA-TGC-TTT-TTT-TGT-TCC-TG
Opo-Env3-Int-F	5'-AAA-TTC-ACC-AAC-AAG-GG
Opo-Env3-Int-R	5'-TAC-AAC-TTC-TGA-GAG-GGA-GGA
Opo-Env2-ORF-F	5'-TAT-GRA-ATC-CYK-CCC-TTC
Opo-Env2-Locus-R	5'-ATA-CAT-ACG-CGT-TAC-AGT-ACT-TGT-TAA-GCA

Table S3. Fusogenic activity of Syncytin-Opo1

Species	Cell	Syncytin-Opo1 cell-cell fusion efficiency					No Env
		<i>M. domestica</i>	<i>M. domestica</i> optimized	<i>M. brevicaudata</i>	<i>M. adusta</i>	<i>M. americana</i>	
Human	293T	-	-	-	-	-	-
	HeLa	-	-	-	-	-	-
	HuH7	-	-	-	-	-	-
	SHSY-5Y	-	-	-	-	-	-
Mouse	MCA205	-	-	-	-	-	-
	WOP	-	-	-	-	-	-
Rat	208F	-	-	-	-	-	-
Hamster	A23	+	++	++	++	++	-
Cat	G355.5	-	-	-	-	-	-
Dog	MDCK	-	-	-	-	-	-
Opossum	OK	-	-	-	-	-	-

Fusogenic activity of the indicated Syncytin-Opo1 on a panel of target cells, including human, rodent, carnivoran, and opossum cells. Same experimental conditions as in Fig. 5. Fusion index was calculated as $[(N - S)/T] \times 100$, where N is the number of nuclei in the syncytia, S is the number of syncytia, and T is the total number of nuclei counted, with the following scale: -, <5; +, 5-10; ++, 10-30.