



**HAL**  
open science

## Genome-Wide Screening of Retroviral Envelope Genes in the Nine-Banded Armadillo (*Dasypus novemcinctus*, *Xenarthra*) Reveals an Unfixed Chimeric Endogenous Betaretrovirus Using the ASCT2 Receptor

Sébastien Malicorne, Cécile Vernochet, Guillaume Cornelis, Baptiste Mulo, Frédéric Delsuc, Odile Heidmann, Thierry Heidmann, Anne Dupressoir

### ► To cite this version:

Sébastien Malicorne, Cécile Vernochet, Guillaume Cornelis, Baptiste Mulo, Frédéric Delsuc, et al.. Genome-Wide Screening of Retroviral Envelope Genes in the Nine-Banded Armadillo (*Dasypus novemcinctus*, *Xenarthra*) Reveals an Unfixed Chimeric Endogenous Betaretrovirus Using the ASCT2 Receptor. *Journal of Virology*, 2016, 90 (18), pp.8132 - 8149. 10.1128/JVI.00483-16 . hal-01879328

**HAL Id: hal-01879328**

**<https://hal.science/hal-01879328v1>**

Submitted on 23 Sep 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Genome-Wide Screening of Retroviral Envelope Genes in the Nine-Banded Armadillo (*Dasypus novemcinctus*, Xenarthra) Reveals an Unfixed Chimeric Endogenous Betaretrovirus Using the ASCT2 Receptor

Sébastien Malicorne,<sup>a</sup> Cécile Vernochet,<sup>a</sup> Guillaume Cornelis,<sup>a,b\*</sup> Baptiste Mulot,<sup>c</sup> Frédéric Delsuc,<sup>d</sup> Odile Heidmann,<sup>a</sup> Thierry Heidmann,<sup>a</sup> Anne Dupressoir<sup>a</sup>

Physiologie et Pathologie Moléculaires des Rétrovirus Endogènes et Infectieux, Unité Mixte de Recherche 9196, Centre National de la Recherche Scientifique, Institut Gustave Roussy, Villejuif, and Université Paris-Sud, Orsay, France<sup>a</sup>; Université Paris Diderot, Sorbonne Paris Cité, Paris, France<sup>b</sup>; ZooParc de Beauval and Beauval Nature, Saint Aignan, France<sup>c</sup>; Institut des Sciences de l'Évolution, UMR 5554, CNRS, IRD, EPHE, Université de Montpellier, Montpellier, France<sup>d</sup>

## ABSTRACT

Retroviruses enter host cells through the interaction of their envelope (Env) protein with a cell surface receptor, which triggers the fusion of viral and cellular membranes. The sodium-dependent neutral amino acid transporter ASCT2 is the common receptor of the large RD114 retrovirus interference group, whose members display frequent *env* recombination events. Germ line retrovirus infections have led to numerous inherited endogenous retroviruses (ERVs) in vertebrate genomes, which provide useful insights into the coevolutionary history of retroviruses and their hosts. Rare ERV-derived genes display conserved viral functions, as illustrated by the fusogenic *syncytin env* genes involved in placentation. Here, we searched for functional *env* genes in the nine-banded armadillo (*Dasypus novemcinctus*) genome and identified *dasy-env1.1*, which clusters with RD114 interference group *env* genes and with two *syncytin* genes sharing ASCT2 receptor usage. Using *ex vivo* pseudotyping and cell-cell fusion assays, we demonstrated that the Dasy-Env1.1 protein is fusogenic and can use both human and armadillo ASCT2s as receptors. This gammaretroviral *env* gene belongs to a provirus with betaretrovirus-like features, suggesting acquisition through recombination. Provirus insertion was found in several *Dasypus* species, where it has not reached fixation, whereas related family members integrated before diversification of the genus *Dasypus* >12 million years ago (Mya). This newly described ERV lineage is potentially useful as a population genetic marker. Our results extend the usage of ASCT2 as a retrovirus receptor to the mammalian clade Xenarthra and suggest that the acquisition of an ASCT2-interacting *env* gene is a major selective force driving the emergence of numerous chimeric viruses in vertebrates.

## IMPORTANCE

Retroviral infection is initiated by the binding of the viral envelope glycoprotein to a host cell receptor(s), triggering membrane fusion. Ancient germ line infections have generated numerous endogenous retroviruses (ERVs) in nearly all vertebrate genomes. Here, we report a previously uncharacterized ERV lineage from the genome of a xenarthran species, the nine-banded armadillo (*Dasypus novemcinctus*). It entered the *Dasypus* genus >12 Mya, with one element being inserted more recently in some *Dasypus* species, where it could serve as a useful marker for population genetics. This element exhibits an *env* gene, acquired by recombination events, with conserved viral fusogenic properties through binding to ASCT2, a receptor used by a wide range of recombinant retroviruses infecting other vertebrate orders. This specifies the ASCT2 transporter as a successful receptor for ERV endogenization and suggests that ASCT2-binding *env* acquisition events have favored the emergence of numerous chimeric viruses in a wide range of species.

Endogenous retroviruses (ERVs) result from occasional infections of a germ line cell by a retrovirus, whose integrated proviral copy is then transmitted vertically in a Mendelian fashion and can become fixed in the host population. After the initial integration, ERVs can amplify themselves and insert into different locations in the genome by either germ line reinfection or intracellular retrotransposition (1), giving rise over long periods of time to several multigenic families of related ERV elements that ultimately make up a significant fraction of vertebrate genomes (8 to 10% of human and mouse genomes) (2–4).

The ability of a retrovirus to enter and replicate in a given host is dependent on various host and viral factors. However, the initial step of infection is essential in mediating entry into the target cell. The viral envelope (Env) glycoprotein is the primary determinant of access and host range, by recognizing a receptor at the surface of target cells and then triggering the fusion of viral and cellular membranes. To date, a large variety of distinct membrane-bound

molecules, usually nutrient transporters, have been identified as retroviral Env receptors (5, 6). However, previous experiments based on viral interference in human cells pointed out the ability of some Env glycoproteins to use common receptors for infection, which led to their classification into interference groups (7). The largest and more widely dispersed group identified within different species, the “RD114 interference group,” consists of type D simian betaretroviruses, i.e., simian retrovirus type 1 (SRV1) to SRV5, Mason-Pfizer monkey virus (MPMV), and squirrel monkey retrovirus (SMRV) (which are prevalent in nonhuman primates and could cause zoonotic infections in humans), as well as three type C retroviruses, i.e., feline RD114 endogenous virus, baboon endogenous virus (BaEV), and avian reticuloendotheliosis virus strain A (REV-A). The common receptor of this interference group has been identified as the sodium-dependent neutral amino acid transporter type 2 (ASCT2), a broadly expressed multipass membrane-bound protein (8).

On the basis of the phylogenetic relatedness of the highly conserved ERV polymerase (*pol*) region to that of infectious retroviruses, ERVs have been separated into three classes: class I, related to gammaretroviruses (or type C retroviruses); class II, related to alpha-, beta-, and deltaretroviruses and to lentiviruses; and class III, distantly related to the spuma-like retroviruses (9, 10). However, phylogenetic trees based on Env sequences result in a different classification for some ERVs (11, 12). For instance, the betaretrovirus genus is typically split between members having a beta-type Env protein and those having a gamma-type Env protein. This can be explained by recombination events involving the acquisition by a retrovirus of a heterologous *env* gene, a process that is supposed to facilitate cross-species transmission (13).

Many ERVs have persisted in the genome of their host for millions of years and are now fixed in most host populations. A large majority of them have been inactivated over time by the accumulation of deleterious mutations or internal recombination events leading to solo long terminal repeat (LTR) formation. A few examples of evolutionarily younger elements whose endogenization is recent or still ongoing and that still segregate in the host population have been described for some mammalian species (14–18). These elements have often maintained intact open reading frames (ORFs) for one or more of their genes and can even be replication competent, as described previously for feline ERVs (17). Some of them have been used as informative genetic markers for investigating host population evolutionary history (18, 19). Besides recently acquired viral ORFs, some vertebrate genomes contain rare copies of ERV-derived genes that are endowed with functions of their viral ancestors and that have been coopted by the host for a physiological role. In mouse, sheep, cat, and chicken genomes, ERV-carried *env* genes can confer resistance to infection by blocking receptors against exogenous retrovirus infection (20). In addition, in various mammalian lineages, the fusogenic properties of ERV *env* genes have been repeatedly subverted to form a fused cell layer in the placenta, the syncytiotrophoblast (21–26; see references 27 and 28 and references therein). These *env* genes, referred to as “*syncytin* genes,” are able to trigger cell-cell fusion upon binding to cell surface receptors. Remarkably, the ASCT2 transporter was identified as the receptor of both human syncytin-1 (22) and rabbit syncytin-Ory1 (29).

A search of the host genome for ERV *env* genes with intact ORFs and endowed with functional properties can thus serve as a useful indication of the evolutionary history, potential host range, and cell tropism of a given retroviral lineage, in addition to allowing the identification of conserved retrovirally derived genes with an impact on host biology, such as the *syncytin* genes (30). Here, we searched for coding *env* genes that would have kept some of the properties of their viral ancestor in an anciently diverged lineage of mammals. We focused on members of the Xenarthra clade that diverged >100 million years ago (Mya) from the three other clades of eutherian mammals in which functional *env* and/or *syncytin* genes have previously been identified (Euarchontoglires, Laurasiatheria, and Afrotheria). We first investigated the nine-banded armadillo (*Dasypus novemcinctus*), whose genome has been sequenced and which belongs to the genus *Dasypus* (long-nosed armadillos; family Dasypodidae, order Cingulata) (31) that contains widely distributed species formerly confined to South and Central America. Remarkably, *D. novemcinctus* was found to have colonized the southern United States during the last 200 years (32). Moreover, the recent low-coverage shotgun genome sequencing of all xenarthran species allowed a robust phylogenetic framework and time scale to be established (31), thus facilitating the dating of retroviral germ line insertions in this mammalian clade.

Here, we report the discovery of an ERV-derived *env* gene in the genome of *D. novemcinctus* that is fusogenic in both pseudotyping and cell-cell fusion assays. We further identify its receptor as the ASCT2 transporter. Interestingly, this gammaretroviral *env* gene belongs to a family of ERVs, which we named DnERV-1, with a betaretrovirus-like genetic organization and phylogenetic relatedness, suggesting acquisition through recombination events. This gene is found in several *Dasypus* species, with evidence of insertional polymorphism, whereas other members of the DnERV-1 family integrated before the diversification of the genus *Dasypus* >12 Mya. This newly described ERV lineage in the *Dasypus* genus, which has not yet reached fixation, constitutes a potential genetic marker to trace the evolutionary history of invasive U.S. populations. This study also extends the conservation of ASCT2 receptor usage to retroviruses infecting members of the clade Xenarthra, a major clade of placental mammals, and suggests that the acquisition of an ASCT2-interacting *env* gene has repeatedly favored the emergence of numerous chimeric viruses, including DnERV-1, in a number of vertebrate species.

## MATERIALS AND METHODS

**Biological material.** Two gravid nine-banded armadillo females from an invasive U.S. population were trapped in Arkansas (Franklin County) in 2012. Two fetuses (ca. 50 g and 10 cm long), placenta, the uterine-placenta section, and amnion were recovered from female 1. The tail and/or liver was dissected from each fetus and recovered for analysis. The whole genital tract, including uterus, vagina, oviduct, and ovary, was recovered from female 2, whose pregnancy was assessed *a posteriori* by detecting the expression of the placenta-specific GCM1 marker gene in this tissue sample (33) (see Fig. 2B). Kidney, spleen, and/or muscle and liver were also collected from each pregnant female. All tissues were immersed in RNAlater, and total RNA was extracted by using the RNeasy RNA isolation kit (Qiagen). Genomic DNA was extracted from the alcohol-preserved livers of the two pregnant females as well as from ear biopsy specimens of other long-nosed armadillo individuals collected in different locations in the United States, Central America (CA), South America, and French Guiana (FG) (Table 1). Genomic DNAs from *Dasypus novemcinctus*

Received 16 March 2016 Accepted 24 June 2016

Accepted manuscript posted online 6 July 2016

**Citation** Malicorne S, Vernochet C, Cornelis G, Mulot B, Delsuc F, Heidmann O, Heidmann T, Dupressoir A. 2016. Genome-wide screening of retroviral envelope genes in the nine-banded armadillo (*Dasypus novemcinctus*, Xenarthra) reveals an unfixed chimeric endogenous betaretrovirus using the ASCT2 receptor. *J Virol* 90:8132–8149. doi:10.1128/JVI.00483-16.

**Editor:** S. R. Ross, University of Illinois at Chicago

Address correspondence to Thierry Heidmann, [heidmann@igr.fr](mailto:heidmann@igr.fr), or Anne Dupressoir, [dupresso@igr.fr](mailto:dupresso@igr.fr).

\* Present address: Guillaume Cornelis, Department of Genetics, Stanford University, Stanford, California, USA.

S.M. and C.V. contributed equally to this work.

This article is contribution ISEM 2016-125 from the Institut des Sciences de Montpellier.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/JVI.00483-16>.

Copyright © 2016, American Society for Microbiology. All Rights Reserved.

TABLE 1 Sample details and status of the DnERV(Env1.1) provirus and related sequences in 55 xenarthran individuals

Species (lineage)	Individual	Origin	Provirus insertion (PCR) <sup>b</sup>	Presence of provirus-related elements (mapping) <sup>c</sup>
<i>Dasypus novemcinctus</i> (U.S.)	T1674	U.S. (MS)	-/-	
<i>Dasypus novemcinctus</i> (U.S.)	T1678	U.S. (TX)	+/- <sup>c</sup>	
<i>Dasypus novemcinctus</i> (U.S.)	BM2	U.S. (TX)	+/- <sup>c</sup>	
<i>Dasypus novemcinctus</i> (U.S.)	T1829	U.S. (LA)	+/- <sup>c</sup>	
<i>Dasypus novemcinctus</i> (U.S.)	JLD128	U.S. (GA)	+/+ <sup>c</sup>	
<i>Dasypus novemcinctus</i> (U.S.)	Female 1	U.S. (AR)	+/- <sup>c</sup>	
<i>Dasypus novemcinctus</i> (U.S.)	Female 2	U.S. (AR)	+/- <sup>c</sup>	
<i>Dasypus novemcinctus</i> (U.S.)	BM1	U.S. (TX)	NA	+
<i>Dasypus novemcinctus</i> (East)	JP1	El Salvador (zoo)	NA <sup>d</sup>	+
<i>Dasypus novemcinctus</i> (East)	JP2	El Salvador (zoo)	+/-	
<i>Dasypus novemcinctus</i> (East)	JP3	El Salvador (zoo)	+/- <sup>c</sup>	
<i>Dasypus novemcinctus</i> (East)	JP4	El Salvador (zoo)	+/- <sup>c</sup>	
<i>Dasypus novemcinctus</i> (East)	JP5	El Salvador (zoo)	+/- <sup>c</sup>	
<i>Dasypus novemcinctus</i> (East)	JP6	El Salvador (zoo)	+/- <sup>c</sup>	
<i>Dasypus novemcinctus</i> (East)	JP7	El Salvador (zoo)	+/- <sup>c</sup>	
<i>Dasypus novemcinctus</i> (East)	JP8	El Salvador (zoo)	+/-	
<i>Dasypus novemcinctus</i> (East)	LS1	Mexico (Chiapas)	-/-	
<i>Dasypus novemcinctus</i> (East)	LS2	Mexico (Chiapas)	-/-	
<i>Dasypus novemcinctus</i> (East)	LS3	Mexico (Oaxaca)	+/-	
<i>Dasypus novemcinctus</i> (East)	LS4	Mexico (Chiapas)	+/-	
<i>Dasypus novemcinctus</i> (East)	NP276	Mexico (Oaxaca)	+/-	
<i>Dasypus novemcinctus</i> (East)	YUC1	Mexico (Yucatan)	-/-	+
<i>Dasypus novemcinctus</i> (East)	YUC2	Mexico (Yucatan)	+/-	
<i>Dasypus novemcinctus</i> (East)	YUC3	Mexico (Yucatan)	-/-	
<i>Dasypus novemcinctus</i> (East)	T2631	Belize	+/- <sup>c</sup>	+
<i>Dasypus novemcinctus</i> (West)	AJR1	Mexico	-/-	
<i>Dasypus novemcinctus</i> (West)	AJR2	Mexico	-/-	
<i>Dasypus novemcinctus</i> (West)	MC151	Mexico (Hidalgo)	NA	+
<i>Dasypus pilosus</i>	MSB 49990	Peru	NA	+
<i>Dasypus pilosus</i>	LSUMZ 21888	Peru	NA	+
" <i>Dasypus sabanicola</i> " <sup>a</sup>	MSB1	Bolivia (Beni)	+/-	
" <i>Dasypus sabanicola</i> " <sup>a</sup>	MSB2	Bolivia (Beni)	+/-	+
" <i>Dasypus sabanicola</i> " <sup>a</sup>	MSB3	Bolivia (La Paz)	-/-	
" <i>Dasypus sabanicola</i> " <sup>a</sup>	NP751	Peru (Cuzco)	+/-	
" <i>Dasypus sabanicola</i> " <sup>a</sup>	T1921	Venezuela (Miranda)	-/-	
" <i>Dasypus sabanicola</i> " <sup>a</sup>	AN1	Bolivia	NA	+
" <i>Dasypus sabanicola</i> " <sup>a</sup>	AN2	Bolivia	+/-	
<i>Dasypus sabanicola</i>	USNM 372834	Venezuela	NA <sup>d</sup>	+
<i>Dasypus novemcinctus</i> FG	T4069	French Guiana	+/-	
<i>Dasypus novemcinctus</i> FG	T4553	French Guiana	+/-	
<i>Dasypus novemcinctus</i> FG	DNO169	French Guiana	-/-	
<i>Dasypus novemcinctus</i> FG	T2863	French Guiana	-/-	
<i>Dasypus novemcinctus</i> FG	T1863	French Guiana	NA	+
<i>Dasypus novemcinctus</i> FG	AP207	French Guiana	NA <sup>d</sup>	+
<i>Dasypus septemcinctus</i>	T3002	Argentina	NA	+
<i>Dasypus hybridus</i>	T3002	French Guiana	-/-	
<i>Dasypus hybridus</i>	ZVC M2010	Uruguay	NA	+
<i>Dasypus kappleri</i>	T4339	French Guiana	-/-	+
<i>Dasypus kappleri</i>	DKA37	French Guiana	-/-	
<i>Chaetophractus villosus</i>	ChaeVill1	France (zoo)	-/-	-
<i>Zaedyus pichiy</i>	T6055	Argentina	-/-	-
<i>Chlamyphorus truncatus</i>	CT1	Argentina	-/-	-
<i>Cabassous unicinctus</i>	T1641	French Guiana	-/-	
<i>Tolypeutes matacus</i>	Tma TB1	U.S. (zoo)	NA	-
<i>Tolypeutes matacus</i>	TolyMat1	France (zoo)	-/-	

<sup>a</sup> Uncertain taxonomy.<sup>b</sup> +/+, diploid provirus insertion; +/-, haploid provirus insertion; -/-, no provirus insertion. NA, not applicable.<sup>c</sup> Full-length *dasy-env1.1* ORF detected at the orthologous locus.<sup>d</sup> Illumina reads mapped to the empty locus.<sup>e</sup> +, Illumina reads mapped to the DnERV(Env1.1) sequence; -, no Illumina reads mapped to the DnERV(Env1.1) sequence.

*tus* FG (DNO169) and *Cabassou uncinatus* were extracted from 95% ethanol-preserved tissues collected in French Guiana and stored in the Mammalian Tissue Collection of the Institut des Sciences de l'Évolution de Montpellier, France (F. Catzeflis). Genomic DNAs from *Chaetophra-tus villosus* and *Tolypeutes matacus* were extracted from blood samples collected by B. Mulot and R. Potier (ZooParc de Beauval and Beauval Nature, Saint Aignan, France) using DNA blood kit II (PaxGene). Genomic DNAs from *Dasyus kappleri*, *Dasyus hybridus*, *Zaedyus pichiy*, and *Chlamyphorus truncatus* were extracted in a previous study (31). Samples used in this study are detailed in Table 1.

**Database screening and sequence analyses.** Retroviral endogenous *env* gene sequences were searched for by using BLAST on the nine-banded armadillo genome (6× coverage assembly of the *D. novemcinctus* genome; Baylor College of Medicine/Dasnov3.0, January 2012). Sequences containing an ORF of >400 codons (from start to stop codons) were extracted from the Dasnov3.0 genomic database using the getorf program of the EMBOSS package (<http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html>) and translated into amino acid sequences. These sequences were blasted against the transmembrane (TM) subunit amino acid sequences of 37 retroviral envelope glycoproteins (from representative ERVs, among which are the known syncytins, and infectious retroviruses) by using the BLASTP program of the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/BLAST>). Putative envelope protein sequences were then selected based on the presence of a hydrophobic domain (transmembrane domain) located 3' to a highly conserved C-X<sub>5,6,7</sub>-C motif. The identified Env-encoding sequence coordinates are listed in Table S1 in the supplemental material.

Multiple alignments of amino acid or nucleotide sequences were carried out by using ClustalW in the SeaView 4 program (34). Maximum likelihood phylogenetic trees were constructed with RAxML version 7.4.2 (35) by using the general time-reversible (GTR) model with the GAMMA distribution of among-site rate variation, with bootstrap percentages being computed after 1,000 replicates.

The nine-banded armadillo genome was secondarily screened with the identified *dasy-env1.1* sequence by using the BLASTN program from the NCBI. Sequences matching (E value of <1e-3) over >70% of the length of the *dasy-env1.1* sequence were extracted, together with 7 kb and 3 kb of upstream and downstream sequences, respectively. After elimination of clearly misassembled elements, consensus sequences were built by using the Multalin program (<http://multalin.toulouse.inra.fr/multalin/>). Finally, DnERV(Env1.1) *gag*- and *pol*-related coding sequences were searched for in the nine-banded armadillo genome by using BLASTN. Sequences displaying >70% identity over >80% of the *gag* or *pol* sequence were extracted, and the putative ORFs were searched by using the getorf program. The NCBI CD-search program was used to find conserved protein motifs.

**Real-time RT-PCR.** Transcript levels were determined by reverse transcription-quantitative PCR (qRT-PCR). Reverse transcription was performed with 500 ng of DNase-treated RNA as described previously (36). PCR was carried out with 5 µl of diluted (1:20) cDNA in a final volume of 25 µl by using Fast SYBR green PCR master mix (Qiagen) with an ABI Prism 7000 sequence detection system. Primer sequences are available upon request. The transcript levels were normalized to the amount of the housekeeping gene *RPL19* (ribosomal protein L19). Samples were assayed in duplicate. The *Dasy-Gcm1* gene, expressed in placental trophoblast cells and widely conserved in mammals (33), was used as a marker to confirm early-stage pregnancy in female 2.

**Search for DnERV(Env1.1) in the genus *Dasyus* and in other xenarthran species.** PCRs were performed on 100 ng of genomic DNA by using Accuprime *Taq* DNA polymerase (Invitrogen). A highly sensitive touchdown PCR protocol was performed (30-s to 3-min elongation time at 68°C, with 30 s of hybridization at temperatures ranging from 60°C to 50°C and -1°C per cycle for 10 cycles, followed by 40 cycles at 50°C). The positions of the primers are shown in Fig. 9B. PCR products were either directly sequenced without cloning to avoid low-level mutations intro-

duced by PCR or cloned into the pGEM-T Easy vector (Promega), followed by sequencing of several individual clones.

Illumina reads previously produced from low-coverage shotgun genome sequencing of all xenarthran species (31) were also used to test for the presence or absence of DnERV(Env1.1)-related copies. Single reads were mapped onto the sequence of the empty locus, the 5'-LTR and 3'-LTR junctions, and the *dasy-Env1.1* ORF by using the medium-sensitivity parameters of the Geneious R9 Mapper (37).

***Dasy-env* expression vectors and cell fusion and pseudotyping assays.** *Dasy-env* ORF fragments were PCR amplified from the genomic DNA of a pregnant *D. novemcinctus* U.S. female with primers defined upstream of the *env* ORF within the provirus and downstream of either the *env* ORF for *dasy-env1.4* ("full *env*" amplicon) (see Fig. 9B) or the putative 3' LTR of the provirus within the flanking region for *dasy-env1.1* ("*env1.1* locus" amplicon) (see Fig. 9B) and cloned into the phCMV-G vector (GenBank accession number AJ318514) (gift from F.-L. Cosset). Cell-cell fusion assays were performed by cotransfection of cells (1 × 10<sup>5</sup> to 2 × 10<sup>5</sup> cells per well in a 6-well plate) with *env*-expressing vectors along with a vector expressing an *nls-lacZ* gene (0.5 to 1 µg at a ratio of 1:1) by using the Lipofectamine LTX reagent transfection kit (Invitrogen). At 24 to 48 h posttransfection, cells were fixed and stained with X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside).

For the pseudotyping assay, *dasy-env*-murine leukemia virus (MLV) pseudotypes were produced by cotransfection of 8 × 10<sup>5</sup> 293T cells in a 60-mm dish with 2.25 µg of pGagPol MLV (encoding MLV retroviral proteins except Env), 2.25 µg of MFG-nlsLacZ (a LacZ-marked defective MLV retroviral vector), and 0.5 µg of *env* expression vectors by using the Lipofectamine LTX transfection kit (Invitrogen). Supernatants from the transfected cells were harvested 48 h after transfection, filtered through 0.45-µm-pore-size polyvinylidene difluoride (PVDF) membranes, supplemented with Polybrene (8 µg/ml), and transferred to target cells seeded into 24-well plates (5 × 10<sup>4</sup> to 8 × 10<sup>4</sup> cells per well) the day before infection, followed by spinoculation at 1,200 × g for 2 h 30 min at room temperature. X-Gal staining was performed at 3 days postinfection.

For cell-cell fusion by the coculture assay, A23 cells were seeded at 5 × 10<sup>5</sup> cells per 60-mm dish. A set of dishes was transfected by using the Lipofectamine LTX kit (Invitrogen) with 5 µg of either a nine-banded armadillo *ASCT2* (see below), a human *ASCT2* (22), or a human *MFSD2* (major facilitator superfamily domain-containing protein 2) (38) expression vector or an empty vector, and another set was transfected with 2.5 µg of either a *dasy-env1.1*, *syncytin-1*, or *syncytin-2* expression vector or an empty vector, each cotransfected with 2.5 µg of the *nls-lacZ* expression vector. One day after transfection, 3.5 × 10<sup>5</sup> cells from each group of transfected cells were cocultured in 6-well plates. Syncytia were visualized by X-Gal staining 24 to 48 h after coculture. The nine-banded armadillo *ASCT2* cDNA was amplified from placenta RNA by RT-PCR using primers encompassing the predicted ORF and SeqAmp DNA polymerase (Clontech) optimized for GC-rich regions and cloned into the phCMV-G vector.

All cell lines were described previously (24, 39). These cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (vol/vol) fetal calf serum (FCS) (Invitrogen), 100 mg/ml streptomycin, and 100 U/ml penicillin.

**Accession number(s).** The newly determined sequence was deposited in GenBank under accession number [KU523789](https://doi.org/10.1093/nucleic/acc123).

## RESULTS

***In silico* search for retroviral *env* genes within the nine-banded armadillo genome.** To identify putative *env*-derived genes in the nine-banded armadillo genome sequence (6× coverage assembly of the *Dasyus novemcinctus* genome; NCBI, Dasnov3.0, January 2012), we made use of a method that we previously devised to screen the tenrec genome for such genes (40). Basically, a BLAST search for ORFs (from the Met start codon to the stop codon) of >400 amino acids (aa) was performed by using a selected series of

Env protein sequences, including all presently identified syncytins (see Materials and Methods). This resulted in a series of sequences that were further selected for the presence of a hydrophobic domain of >20 aa located 3' to a C-X<sub>5,6,7</sub>-C motif, corresponding to highly conserved motifs of retroviral envelope proteins (C-C and transmembrane domains) (Fig. 1A). This yielded 162 sequences, which were incorporated into an Env phylogenetic tree, including representatives of the known retroviral classes, a series of ERV Env proteins, as well as the known syncytins (Fig. 1B). A large fraction of these sequences, i.e., 140, could be assigned to class I ERVs related to exogenous gammaretroviruses (e.g., Moloney murine leukemia virus [MoMLV]), whereas 22 are homologous to class II ERVs related to exogenous betaretroviruses (e.g., murine mammary tumor virus [MMTV]). Using their phylogenetic proximity and/or the ability to design PCR primers common to most or all sequences as a criterion (see below), some sequences were grouped into single families, resulting ultimately in 14 Env protein families that we named Dasy-Env1 to -Env14 (Fig. 1B). Several members of these *env* coding gene families are present at very high copy numbers (up to 35 copies), suggesting recent amplification bursts. They were most probably formed by independent virus integration or retrotransposition events but not by genome duplication, since no homology was found between the *env*-containing proviral flanking sequences (data not shown).

*In silico* analysis of the overall structure of the 14 identified *dasy-env* gene families (Fig. 1C) strongly suggests that they encode bona fide retroviral Env proteins, with recognizable characteristic features including the presence of a standard furin cleavage site delineating a surface (SU) subunit and a transmembrane (TM) subunit (consensus motif of R-X-K/R-R but with a degenerate sequence in some family members) and a CXXC motif in the SU subunit involved in SU-TM interactions (class I Env). Hydrophobicity plots identify the hydrophobic transmembrane domain within the TM subunit required for anchoring of the Env protein within the plasma membrane and a putative hydrophobic fusion peptide at the TM subunit N terminus. Class I Envs contain a canonical immunosuppressive domain (ISD). In some cases, a clear signal peptide at the N terminus of the sequences was predicted by the Phobius program (<http://phobius.sbc.su.se/>).

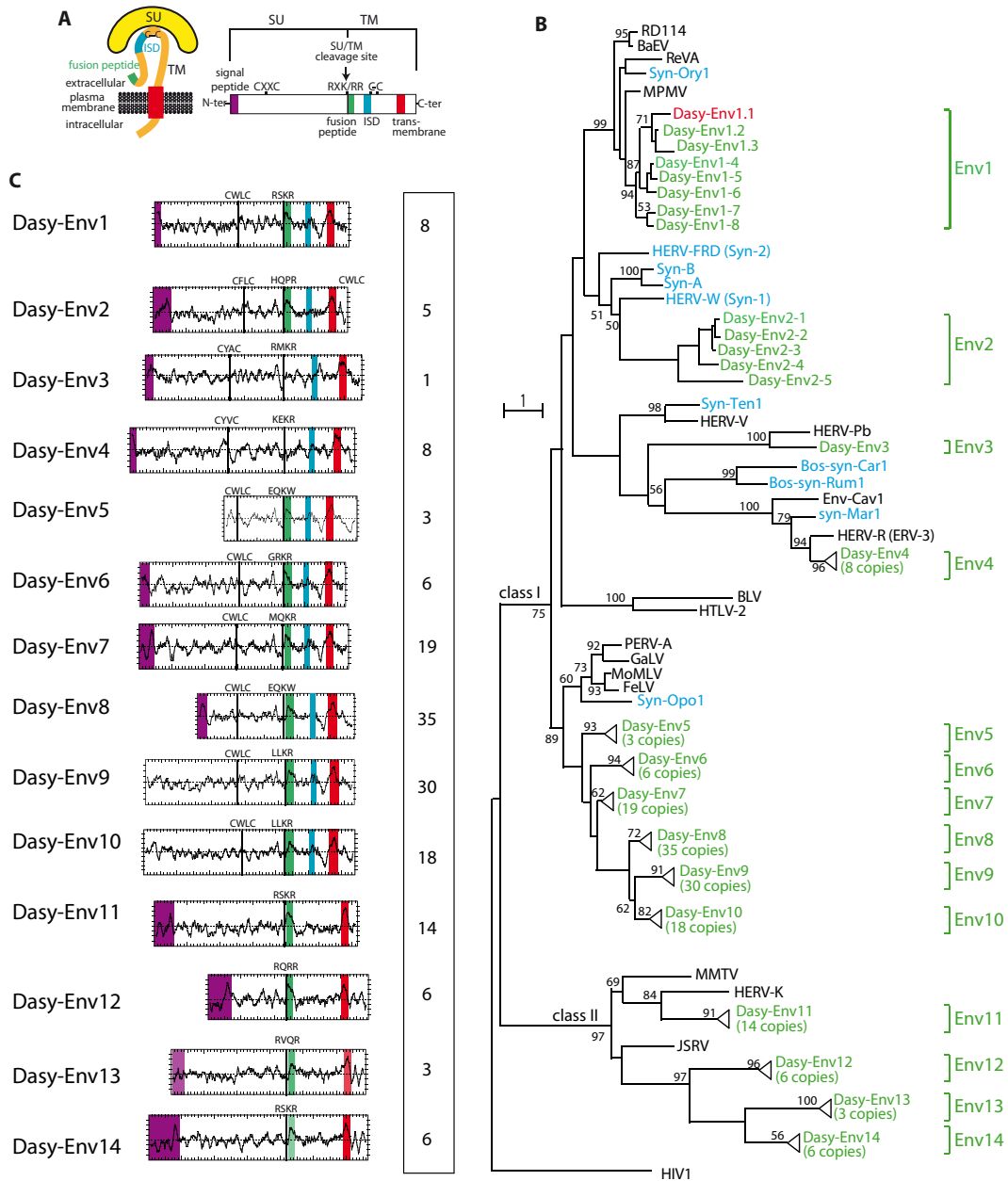
**Search for candidate functional retroviral *env* genes.** Quantitative RT-PCR analysis, performed by using primers that were designed to match all *env* sequences within each family of elements (see Materials and Methods) and a series of tissues, including placenta tissue (recovered from two pregnant armadillo females at different gestational stages), disclosed expression for most *env* gene families, with various transcript levels among the selected tissues (Fig. 2A). Of note, none of them showed one of the characteristic properties of *syncytin* genes, i.e., at least 3- to 4-fold-higher expression levels in the placenta than in other tissues. However, the inability to target each candidate *env* gene individually (due to high *env* copy numbers and/or strong sequence homology within most *env* families) precluded the determination by RT-PCR of the exact contribution of each *env* gene to this expression profile and, therefore, an unambiguous identification of candidate *syncytin* genes based on this criterion.

A phylogenetic proximity-based criterion was then used to select candidate functional *env* genes. Several endogenous and exogenous retroviral *env* genes, among which are the human *syncytin-1* and the rabbit *syncytin-Ory1* genes, as well as the RD114 interference group of retroviruses (including RD114, MPMV,

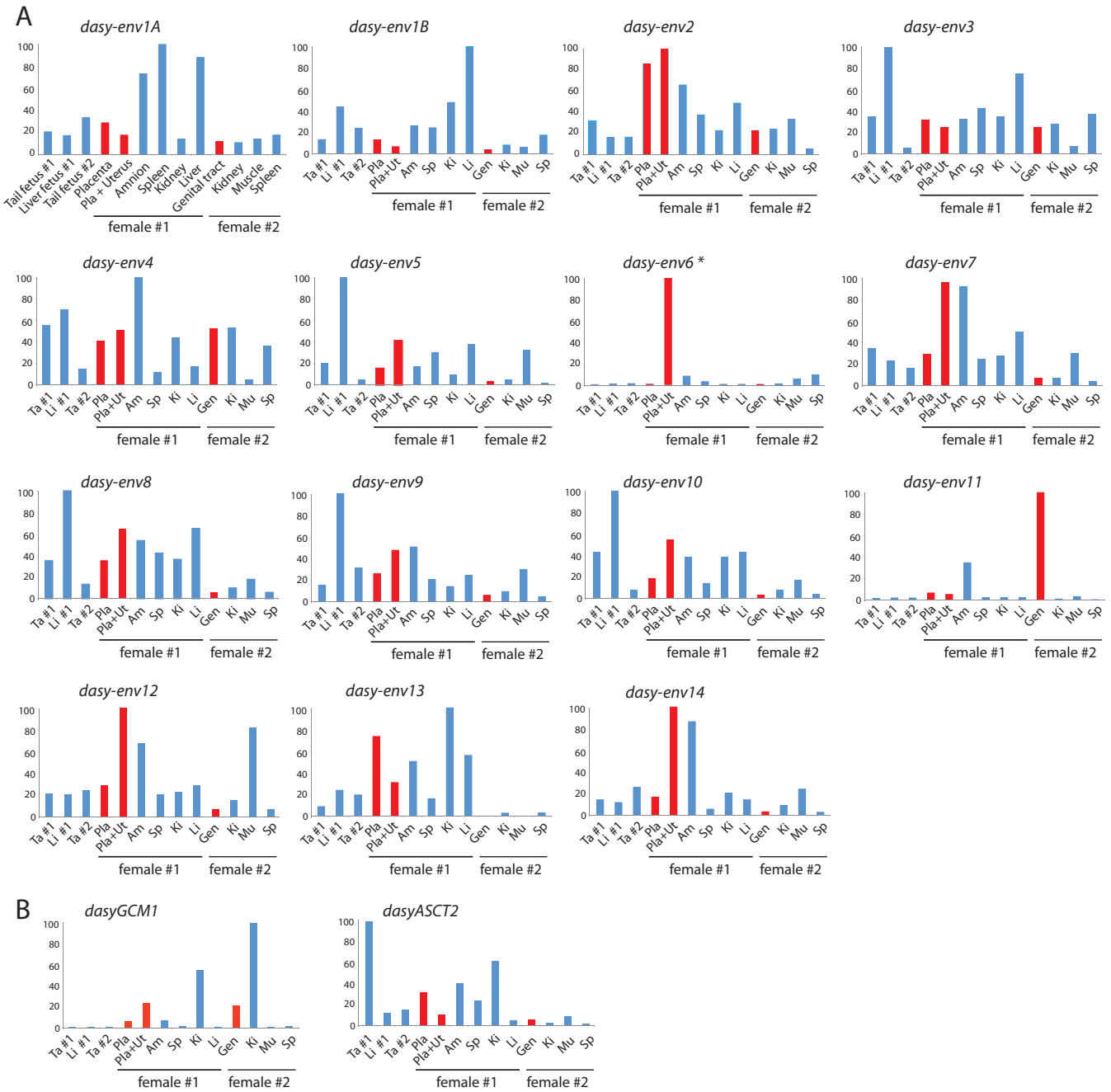
BaEV, and REV-A) were previously reported to use the sodium-dependent neutral amino acid transporter ASCT2 as a cell surface receptor. Interestingly, as shown in Fig. 1B, the members of the *dasy-env1* gene family in the armadillo genome clustered with the above-mentioned *syncytin* and *env* genes. This makes them candidate genes encoding an Env protein that uses ASCT2 as a receptor. Among them, only two representatives, Dasy-Env1.1 and Dasy-Env1.4, harbor all the critical determinants of a functional Env protein, i.e., a predicted signal peptide at the SU subunit N terminus, a consensus furin cleavage site (R-X-K/R-R), and the presence of a hydrophobic fusion peptide at the TM subunit N terminus (Fig. 3). Both proteins were tested further for their functionality.

**Dasy-Env1.1 is a fusogenic retroviral envelope protein that interacts with the ASCT2 transporter.** The functionality of Dasy-Env1.1 and Dasy-Env1.4 as retrovirally derived, fusogenic Env proteins was investigated *ex vivo* by testing their ability to render a recombinant retrovirus, deprived of its own *env* gene, able to infect target cells through virus-cell membrane fusion. To do so, the *dasy-env1.1* and *dasy-env1.4* sequences were PCR amplified from armadillo genomic DNA and cloned into a cytomegalovirus (CMV) promoter-containing expression vector (see Materials and Methods), and the plasmids containing the corresponding *env* genes (100% identical to the genomic sequence) were assayed. Pseudotypes generated in human 293T cells with Dasy-Env1.1 Env and an MLV core were able to infect a large panel of cells of different origins, including human (SH-SY5Y, HeLa, 293T, and TE671), simian (Vero), rodent (208F and MCA205), and carnivore (G355-5 and MDCK) cells, with the exception of Chinese hamster A23 cells (illustrated in Fig. 4A for HeLa, SH-SY5Y, and A23 cells). In contrast, the Dasy-Env1.4 protein did not mediate infection of these cells, as assayed in parallel in the same experiments (not shown). The fusogenic properties of Dasy-Env1.1 and Dasy-Env1.4 were further tested in a cell-cell fusion assay involving the transfection of cells with expression vectors for each *env* gene and detection of syncytium formation at 48 h posttransfection, as described above for *syncytin* genes. As shown in Fig. 4B, transient transfection of human HeLa and SH-SY5Y cells with the Dasy-Env1.1-expressing vector (supplemented with a  $\beta$ -galactosidase expression vector) led to the formation of multinucleate LacZ-positive (LacZ<sup>+</sup>) syncytia. Cell-cell fusion was not observed with any of the other cells found to be positive in the pseudotyping experiments (data not shown) (a result similar to that obtained with *syncytin-Ory1* [data not shown] and which indicates different requirements for virus-cell and cell-cell fusion processes). No cell-cell fusion occurred with A23 cells, which could therefore be used for the identification of the Dasy-Env1.1 receptor (see below). Again, no cell-cell fusion was observed for Dasy-Env1.4 assayed under the same conditions (not shown), and this gene was therefore not investigated further.

Given the phylogenetic relationships of Dasy-Env1.1 with ASCT2-interacting Env proteins (see above), we further tested whether Dasy-Env1.1 could possibly use the ASCT2 transporter as a receptor by performing *ex vivo* cell-cell fusion assays. An ORF displaying 77.8% amino acid identity to the human ASCT2/SLC1A5 transporter (hASCT2) was RT-PCR amplified from armadillo RNA (here named dasyASCT2 [GenBank accession number KU523789]) and cloned into an expression vector. In the experiments illustrated in Fig. 5, distinct pools of A23 cells were transfected with an hASCT2, dasyASCT2, or Dasy-Env1.1 expres-



**FIG 1** Characterization of the identified *Dasyus novemcinctus* candidate *env* genes. (A) Schematic representation of a retroviral Env protein delineating the surface (SU) and transmembrane (TM) subunits. The furin cleavage site (consensus, R-X-R/K-R) between the two subunits, the C-X-X-C motif involved in the SU-TM interaction, the hydrophobic signal peptide (purple), the fusion peptide (green), the transmembrane domain (red), and the putative immunosuppressive domain (ISD) (blue) along with the conserved C-X<sub>5,6,7</sub>-C motif are indicated. (B) Retroviral envelope protein-based phylogenetic tree with the newly identified Dasy-Env protein candidates and families. Shown is a maximum likelihood tree using TM subunit amino acid sequences from a series of endogenous and infectious retroviruses and the syncytins. The horizontal branch length is proportional to the percentage of amino acid substitutions from the node (bar on the left), and the percent bootstrap values (>50%) obtained from 1,000 replicates are indicated at the nodes. Green, *Dasyus novemcinctus* candidate *env* genes; red, functional *dasy-env1.1*; blue, syncytin genes ("syn") so far identified; black, selected series of infectious or endogenous retroviral *env* genes. RevA, avian reticuloendotheliosis virus; MPMV, Mason-Pfizer monkey virus; Env-Cav1, "syncytin-like" *Cavia porcellus* Env1 protein; GaLV, gibbon ape leukemia virus; PERV, porcine endogenous retrovirus; BaEV, baboon endogenous virus; HIV1, human immunodeficiency virus type 1; HERV, human endogenous retrovirus; MoMLV, Moloney murine leukemia virus; HTLV-2, human T-lymphotropic virus type 2; BLV, bovine leukemia virus; JSRV, Jaagsiekte retrovirus; MMTV, murine mammary tumor virus; FeLV, feline leukemia virus; RD114, feline endogenous type C retrovirus. (C) Characterization of the candidate *Dasyus* Env proteins. (Left) Hydrophobicity profile for representative candidates within each family is shown with the canonical structural features highlighted in panel A positioned, when present (same color code). Genomic coordinates for the Env protein-encoding sequences are given in Table S1 in the supplemental material, together with the identity of the representative Env candidates. (Right) Number of full-length *env* gene ORFs within each family of elements.



**FIG 2** Real-time qRT-PCR analysis of the candidate *env*, *GCM1*, and *ASCT2* gene transcripts from *Dasybus novemcinctus* placenta and other tissues. Transcript levels are expressed as a percentage of the maximum and were normalized to the amount of *RPL19* control gene RNA (see Materials and Methods). The placenta, placenta plus uterus, whole genital tract (including uterus, vagina, oviduct, and ovaries), fetal tissues, and somatic tissues were obtained from two *D. novemcinctus* pregnant females (see Materials and Methods), as indicated (female 1, late gestation; female 2, early gestation). Tissues are displayed in the same order in all panels (tissues are abbreviated in all but one panel). Red bars correspond to placenta-containing tissues. (A) Results obtained for the 14 *env* gene families. Primers were designed to amplify all copies within each family, except for the *Dasy-Env1* family, for which 2 primer pairs (A and B) should be designed to fit all elements (primer sequences are available upon request). The asterisk indicates the detection of *dasy-env6* transcripts in the mixed placenta-uterus sample but not in the placenta alone, which suggests that the expression of this gene family is restricted to the uterus. (B) Results obtained for *dasyGCM1* and *dasyASCT2*. *GCM1* is a well-conserved transcription factor-encoding gene, which displays specific expression in the placenta and in adult kidney and thymus in some species (33). The presence of *dasyGCM1* transcripts in the whole genital tract of female 2 confirms its pregnancy.

sion vector supplemented with a  $\beta$ -galactosidase expression vector. The *env*-transfected cells were mixed with each of the two receptor-transfected cell populations and assayed for cell-cell fusion. Cell-cell fusion (as revealed by the presence of large LacZ<sup>+</sup>

syncytia) could be observed with both the *Dasy-Env1.1/dasyASCT2* and the *Dasy-Env1.1/hASCT2* pairs. Cell-cell fusion was also observed in parallel with the human syncytin-1/*dasyASCT2* pair (as well as with the syncytin-1/*hASCT2* and syn-





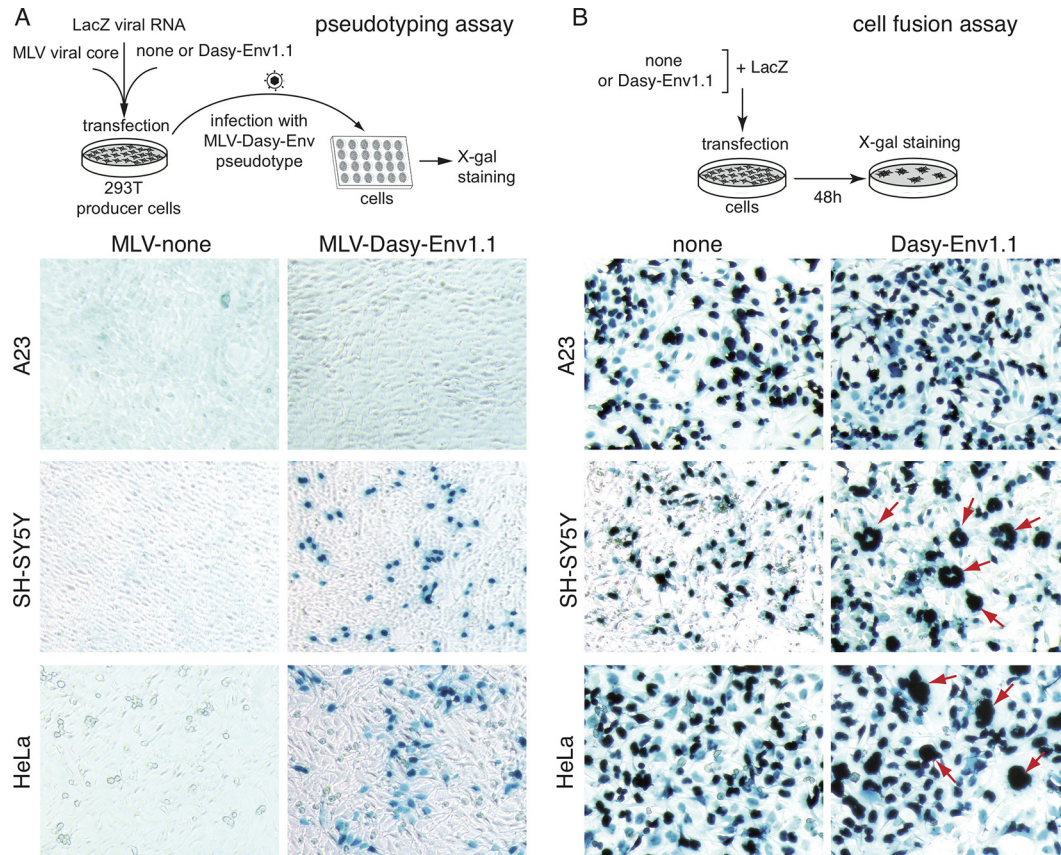
**FIG 3** Primary amino acid sequence and characteristic structural features of the Dasy-Env1.1 and Dasy-Env1.4 retroviral envelope proteins. The SU and TM subunits are delineated, and the furin cleavage site (consensus, R-X-R/K-R) between the two subunits together with the CWLC and CX6CC domains involved in the SU-TM interaction are indicated. The hydrophobic signal peptide, the fusion peptide, the transmembrane domain, and the putative immunosuppressive domain (ISD) are also indicated. The canonical SDGGGX2DX2R motif involved in syncytin-1/hASCT2 interactions is shaded in gray. Of note, this motif is degenerate in Dasy-Env1.4, consistent with the lack of fusogenicity of this gene (see Results).

cytin-2/MFSD2 pairs [38], used as positive controls) but not with any other combinations. These results provide evidence that Dasy-Env1.1 uses ASCT2 as a cell surface receptor and that both human and nine-banded armadillo ASCT2s can function as receptors. Dasy-Env1.1 can therefore be considered a functional fusogenic Env protein that uses ASCT2 as a receptor in the nine-banded armadillo.

**Characterization of the *dasy-env1.1*-containing provirus and of related sequences in the nine-banded armadillo genome.** Examination of the *dasy-env1.1* gene shows that it is part of a proviral structure, here designated DnERV(Env1.1), for *Dasyus novemcinctus* endogenous retrovirus containing Env1.1 (Fig. 6A). It contains intact *pro*, *pol*, and *env* genes but a defective *gag* gene with several frameshift mutations. Moreover, the *pol* gene displays homology to an accessory ORF of Jaagsiekte sheep retroviruses (*orf-x*; conserved protein domain family cl04426) with an as-yet-unknown function. DnERV(Env1.1) has typical type D betaretroviral structural characteristics: its genetic organization requires frameshifts at the *gag-pro* and *pro-pol* overlaps, it encodes a dUTPase (cd07557) at the 5' end of its *pro* gene (41), and a primer binding site (PBS) sequence complementary to armadillo tRNA<sup>Lys1,2</sup> can be identified downstream of the 5' LTR (42). Canonical donor and acceptor splice sites are detected downstream of the 5' LTR and upstream of the Env initiation codon, respectively, as classically observed for retroviruses. The viral core elements are flanked by two short LTRs of 484 nucleotides. LTR pairs differ by 5 nucleotides (out of 484), as assessed by sequencing of the 5' and 3' LTRs from several individuals that were heterozygous for the provirus (see below). Since LTRs are identical at the time of integration, and by applying an estimated neutral substi-

tion rate for the xenarthran genomes of  $1.83 \times 10^{-9}$  substitutions per site per year (43), we calculated an average estimated integration time of 5.5 Mya for this provirus (although this estimation could be altered by possible gene conversion or LTR recombination events that might have occurred since integration). The provirus flanking sequences display a conserved target site duplication (TSD) of 6 nucleotides (Fig. 6A).

With the aim to identify DnERV(Env1.1)-related sequences in the nine-banded armadillo genome to further characterize their *env* and *pol* features (see below), we first screened the genome assembly for the presence of high-quality *dasy-env1.1*-related sequences and then examined their associated proviral structures. Forty-seven *env* genes with homology to over 70% of *dasy-env1.1* sequence were found. These genes could be further grouped into three distinct phylogenetic subfamilies (Fig. 6B), including non-coding copies as well as the eight previously identified Env-encoding genes belonging to the Dasy-Env1 family (Fig. 1). A consensus ORF sequence generated from the alignment of all *env* genes displays the characteristic features of gammaretroviral class I Env proteins, with the CWLC motif in the SU subunit, which has been shown to be involved in MLV SU-TM interactions; a recognizable furin recognition and cleavage site separating the SU and TM domains; the immunosuppressive domain (ISD); and a CX6CC motif typical of the gammaretroviral (and alpharetroviral), but not the betaretroviral, TM protein (44) (data not shown). Examination of the regions surrounding the *env* sequences identified, for most of them, proviral structures with putative *gag*, *pro*, *pol*, and *env* regions flanked by an LTR. A consensus sequence was generated from an alignment of the sequences within each subfamily (Fig. 6C). Interestingly, all proviruses share the same be-

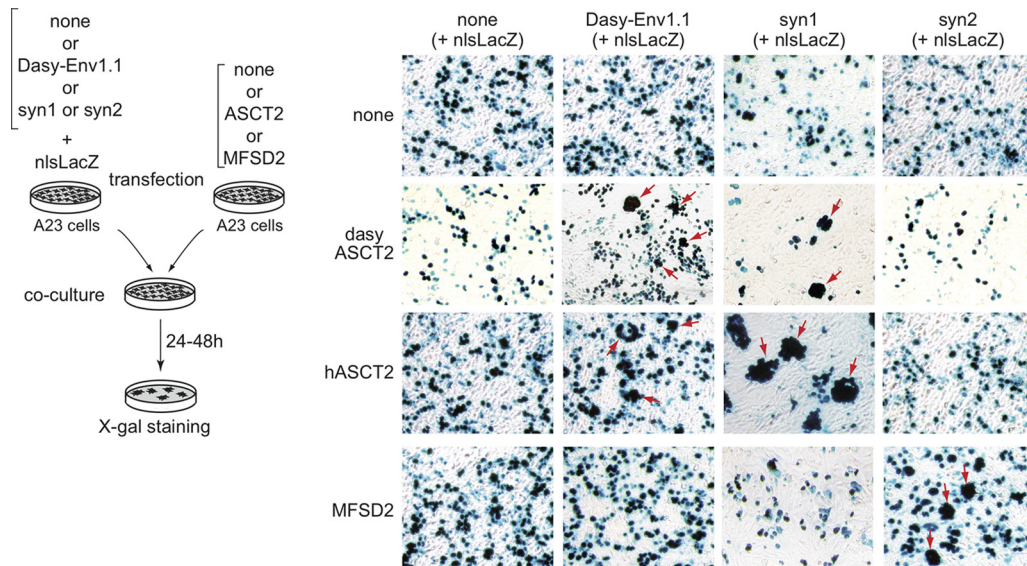


**FIG 4** Dasy-Env1.1 is a fusogenic retroviral envelope protein. (A) Assay for cell infection mediated by Dasy-Env1.1-pseudotyped virus particles. Pseudotypes were produced by cotransfection of human 293T cells with expression vectors for MLV core, the Dasy-Env1.1 protein (or an empty vector), and a *lacZ*-containing retroviral RNA. Supernatants were used to infect a panel of target cells, which were stained with X-Gal 3 days after infection. Infected *LacZ*<sup>+</sup> cells were detected in a series of cell lines (see Results), including human HeLa and SH-SY5Y cells, but not in A23 cells. (B) Assay for cell-cell fusion mediated by Dasy-Env1.1. A panel of cell lines was transfected with an expression vector for Dasy-Env1.1 or an empty vector (none) together with a *LacZ* expression vector. Cells were cultured for 48 h after transfection, fixed, and stained with X-Gal. *LacZ*<sup>+</sup> syncytia (red arrows) were detected in human HeLa and SH-SY5Y cells but not in A23 cells, with only mononucleated cells being visible by using the empty vector. Original magnification,  $\times 100$ .

retroviral genomic organization as that of the DnERV(Env1.1) provirus, e.g., the presence of a *pro* N-terminal dUTPase, strongly suggesting that they in fact originate from a single integration event and belong to a unique multicopy family (“DnERV-1”) present at a high copy number in the nine-banded armadillo genome. Of note, none of the *env*-containing elements displays intact *gag-pol* ORF sequences. In addition, BLAST searches using the *gag* and *pol* sequences of DnERV(Env1.1) as a query revealed only 1 and 4 genes with full coding capacities, respectively, in the nine-banded armadillo genome, with none of them being associated with the same provirus. Finally, a search with less stringent criteria uncovered numerous truncated and/or degenerate proviral sequences, suggesting a long period of activity of the DnERV-1 family in the nine-banded armadillo genome (see below).

**DnERV *env* but not *pol* is of the gamma type, with conservation of ASCT2-binding domains.** Tblastx searches against the NCBI nonredundant sequence database revealed significant amino acid identity (45 to 48%) of Dasy-Env1.1 to the Env proteins from retroviral members of the RD114 interference group (SRV, MPMV, SMRV, RD114, BaEV, and REV-A) and from ERVs identified in bat (*Desmodus rotundus* endogenous betaretrovirus [DrERV]) (45), possum (*Trichosurus vulpecula* endogenous retro-

virus [TvERV(D)]) (46), and mouse (*Mus musculus* endogenous retrovirus MMERV- $\beta 6$ \_NT\_039167) (47) as well as to rabbit *syncytin-Ory1* and human *syncytin-1*. Interestingly, an alignment of the Env SU subunits from all these retroviral sequences identified in a wide range of vertebrate species displayed strong conservation of the SDGGGX2DX2R motif proven to be directly involved in the syncytin-1/ASCT2 interaction as well as of the three cysteine-containing motifs (PCXC, CYX5C, and CX7-9CW) that define structurally conserved SU subdomains in syncytin-1 (48) (Fig. 7). This suggests that ASCT2 usage is, or has been, shared by all these endogenous and infectious retroviruses, as demonstrated for human and rabbit syncytins, for RD114 interference group retroviruses and DnERV(Env1.1) (this study), and possibly, as we now propose, for the Env proteins from chimeric bat DrERV, possum TvERV(D), and murine MMERV- $\beta 6$ . As illustrated in Fig. 8, a phylogenetic analysis based on the highly conserved TM domain of the corresponding Env sequences together with representatives of mammalian retroviral genera indicated that Dasy-Env1.1 and its related sequences group with the Env proteins from typical type C gammaretroviruses (e.g., MoMLV, feline leukemia virus [FeLV], and porcine endogenous retrovirus A [PERV-A]). In contrast, a different phylogenetic pattern is observed for the *pol*-de-



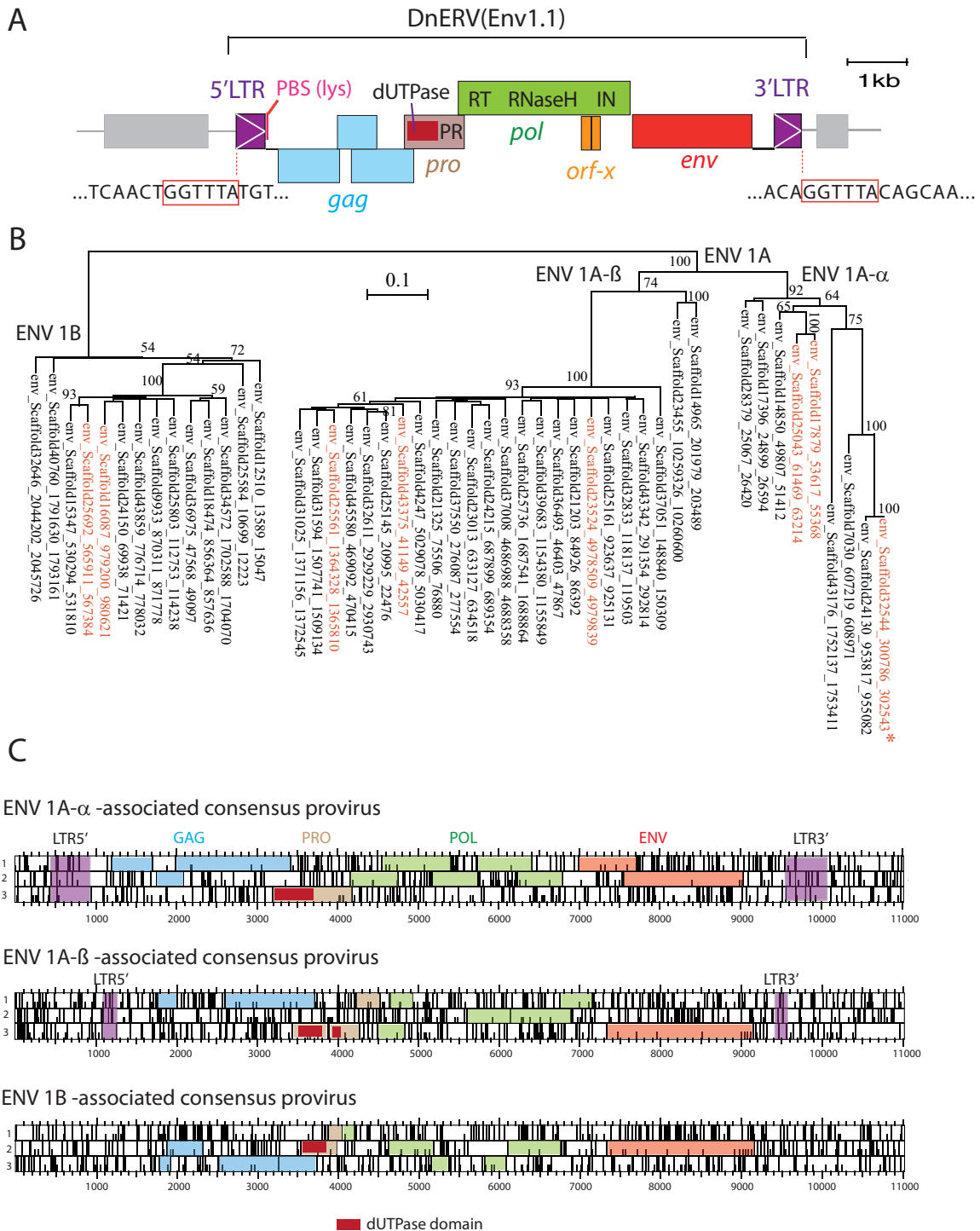
**FIG 5** Fusion assay of ASCT2- and Dasy-Env1.1-transfected cocultured cells demonstrates that ASCT2 is the Dasy-Env1.1 receptor. Cell-cell fusion was assayed upon independent transfections of a set of A23 cells with an empty vector (none) or an expression vector for either the Dasy-Env1.1, syncytin-1 (syn1), or syncytin-2 (syn2) protein together with an *nls-lacZ* gene expression vector and another set of A23 cells with an expression vector for human ASCT2 (hASCT2), *Dasyus* ASCT2 (dasyASCT2), the human syncytin-2 receptor MFSD2 (38), or an empty vector (none). One day after transfection, cells were resuspended, and pairs of transfected cells from each set were cocultured for 1 to 2 days, fixed, and stained with X-Gal. Syncytia can be easily detected (arrows) for the Dasy-Env1.1/dasyASCT2, Dasy-Env1.1/hASCT2, syncytin-1/hASCT2, syncytin-1/dasyASCT2, and syncytin-2/MFSD2 pairs, with only mononucleated cells being visible in the other cases.

rived tree. In fact, consistent with its genomic characteristics, the DnERV(Env1.1) *pol* sequence grouped with those of betaretroviruses (e.g., human endogenous retrovirus K [HERV-K], MMTV, and endogenous Jaagsiekte sheep retrovirus [enJSRV]), as also observed for most of the ERVs harboring a gamma-type Env as described above [i.e., TvERV(D), SMRV, DrERV, syncytin-Ory1, SRV, MPMV, and MMERV- $\beta$ 6], indicating a chimeric structure for all of these retroviral elements. The beta/gamma-chimeric nature of simian retroviruses (49), DrERV (45), and TvERV(D) (46) were previously reported. Armadillo DnERV(Env1.1), murine MMERV- $\beta$ 6, and rabbit syncytin-Ory1-ERV are now added to this list. Other betaretroviruses, such as enJSRV and MMTV, did not fall into this category. This chimerism strongly suggests the occurrence of specific recombination events. DnERV(Env1.1), as for the other chimeric retroviruses, most probably independently acquired a gammaretroviral *env* gene, which may confer a selective advantage to the viruses. ASCT2 receptor usage may represent the selective force driving *env* gene recombination events (see Discussion).

**Distribution of the DnERV(Env1.1) provirus and related elements in the genus *Dasyus*.** The *Dasy-env1.1* provirus was found in the genome of an individual belonging to the U.S. *D. novemcinctus* population. This population originated from a few individuals that recently colonized the southern United States from Mexico (32). Two different *D. novemcinctus* mitochondrial lineages (West and East) in Mexico and in other Central American countries have been described (50), with the U.S. population probably being derived from the East lineage (Fig. 9A). As shown in Fig. 9A, the genus *Dasyus* (family Dasypodidae) also includes several other species (all with a South American geographic localization), whose phylogenetic relationships were recently established thanks to low-coverage shotgun genome sequencing (31).

The distribution of the DnERV(Env1.1) provirus in the genus *Dasyus* was assessed by a PCR-based assay using different sets of primers designed to amplify (i) the junction between the 5'-flanking genomic region and the provirus 5' LTR (designated "5' junction") (Fig. 9B), (ii) the junction between the provirus 3' LTR and the 3'-flanking genomic region (designated "3' junction") (Fig. 9B), and (iii) the empty preinsertion locus, or solo LTR locus (designated "empty locus") (Fig. 9B). PCR was carried out under low-stringency conditions, and PCR amplicons were designed to be short (<165 bp) to enable robust amplification of DNA of variable quality as far as possible. DNA loci were considered negative for proviral insertion when no amplification of the 5' and 3' junctions, using two different sets of primers, could be obtained whereas the empty locus could be amplified. Although we cannot formally exclude the possibility that sequences may be too divergent or altered to allow primer annealing and PCR amplification, the use of low-stringency PCR conditions and of primers targeting different regions should minimize this possibility. Together, these primer sets should determine for an individual whether the DnERV(Env1.1) provirus is present in both chromosomes, in one chromosome, or in none of them. Sequence analysis was performed on the PCR fragments to unequivocally identify preintegration loci or the presence of identical target site duplications flanking the proviruses; we did not find any evidence for solo LTR alleles.

PCRs were attempted for 38 individuals across 5 species of the *Dasyus* genus (Table 1). As indicated in Table 1, among the U.S. invasive *D. novemcinctus* population, one individual was found to be homozygous for the provirus, and five were found to be heterozygous, whereas one was found to be negative, indicating polymorphism of the DnERV(Env1.1) proviral locus across the U.S. armadillo population. Among the *D. novemcinctus* popula-



**FIG 6** Characterization of the DnERV(Env1.1) provirus and related sequences. (A) The DnERV(Env1.1) provirus, with the *gag*, *pro*, *pol*, and *env* genes; the identified typical domains (including the characteristic betaretroviral pro-N-terminal dUTPase domain); and the proviral 5' and 3' LTRs indicated. Repeated long interspersed nuclear elements (LINE) retrotransposons (gray), as identified by the RepeatMasker Web program, are positioned. The 6-bp target site duplication flanking the provirus (red boxes), a characteristic feature of retroviral integration, is indicated. An ~400-bp region downstream of the *env* gene (corresponding to part of the 3' LTR) was PCR amplified and sequenced to fill the unresolved gap (a series of N's) in the Dasnov3.0 database. In addition, the 5' and 3' LTRs of the DnERV(Env1.1) provirus were PCR amplified and sequenced from the genomes of 5 U.S. *D. novemcinctus* individuals heterozygous for the proviral insertion. PBS, primer-binding site. (B) Maximum likelihood tree with the TM subunit amino acid sequences of the identified *dasy-env1.1*-related sequences in the armadillo genome. The vertical branch length is proportional to the percentage of amino acid substitutions from the node (bar on the top), and the percent bootstrap values (>50%) obtained from 1,000 replicates are indicated at the nodes. In red are the eight previously identified *env* sequences encoding full-length ORFs (Fig. 1). The asterisk indicates the functional *dasy-env1.1* gene. (C) Consensus sequences generated from an alignment of the provirus sequences containing *dasy-env1.1*-related genes for each of the three phylogenetic groups determined as described above for panel B (ORF maps). The *gag*, *pro*, *pol*, and *env* genes with the characteristic betaretroviral pro-N-terminal dUTPase domain and the proviral 5' and 3' LTRs (when present) are indicated.

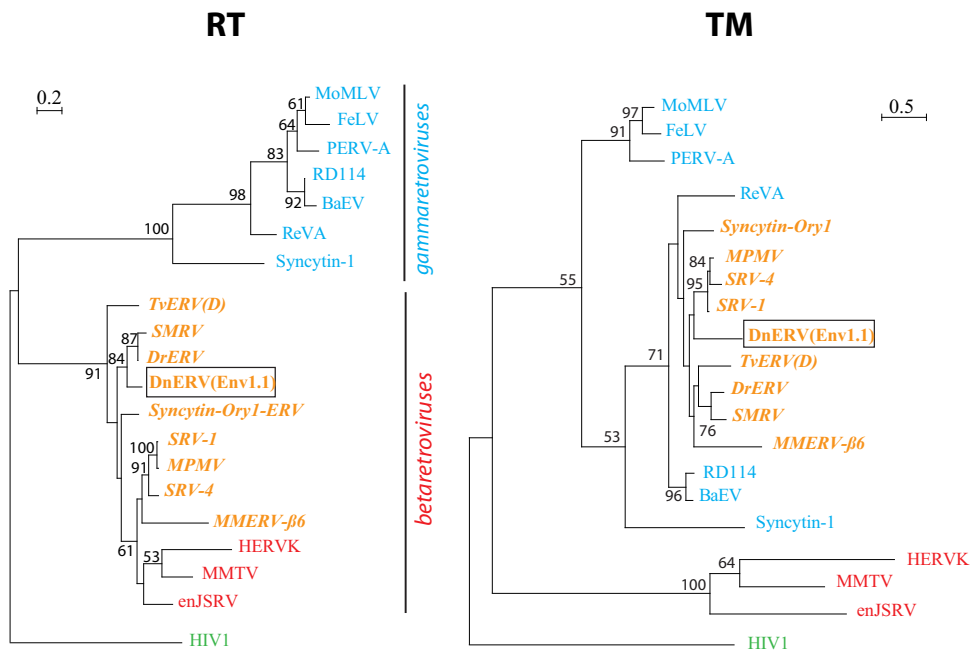
**SU subunit**

human	Syncytin-1	MALP	13	TLTAPP	PCRCMTS	21	RSLSKGTPTF	5	MPRNCYHSATLCMHANTHYWTGKM	1	NPSCP	PG-GLGVTV	CWTFYFTQTGM	SDGGG	VQ	QAREKHVKEVI	180	RNKR	
armadillo	Dasy-Env1.1	MHSF	29	KTLYGT	PCERGG	36	QWVCRNLPRI	19	MHVHCYSEVQSTRNKNKTYTAIL	25	QAGCYG-TIGD	TVCNKIPPAHISDGGG	PQ	QARQII	VQERM	184	RSKR		
rabbit	Syncytin-Ory1	MALS	31	QQH--	HPCTCSGG	35	RWFCVVPKPKL	20	VHSSCYTQYQTCNKGQRLLFAVI	18	SPSCP--	TRGQVACWYPIAPVGVSDGGG	PQ	TLKLTQ	KVVL	187	RQKR		
simian	SRV-1	MNFN	30	QKHGKPCDC	CAGG	35	KWQCVSTPTT	20	VHATCYNHVYQCTIGNKTYLTATM	24	IAGCPENKKGQVVCWNSQPSVHMSDGGG	PQ	KVHEII	VNKKF	197	RPKR			
simian	SRV-4	MNFE	30	QKHGKPCDC	CKG	35	KWQCVSTPKT	20	VHSTCYSSYQCTIGNKTYFTATI	24	SAGCPKDELGKTA	CWSATPPVHVSDGGG	PQ	KVHEII	VQKFF	192	RTKR		
simian	MPMV	MNFN	30	QKHGKPCDC	CAGG	35	KWQCVSTPTT	20	VHSSCYNHVYQCCNIGNKTYLTATI	24	TAGCPNGKKGQVVCWNSRPSVHISDGGG	PQ	KARDII	VNKKF	196	KAKR			
simian	SMRV	MLCI	25	TALFGAPDC	CKGG	36	KWLCVVRPPS	22	MHSSCYSSFSQCTQGNNTYFTAIL	22	QAGCDG-TVGKSV	CWQQAPIHVS	SDGGG	PQ	AVRELY	VQKQI	185	RQKR	
simian	BAEV	MGFT	26	QKRYGRPCDC	CSGG	35	RWKCKSIPKD	18	VHSSCYTSYQCRSGNKTYFTATL	21	QSPC-NGIRGQSI	CWSTTAPIHVS	SDGGG	PL	TRIKSV	QRKL	190	RPKR	
cat	RD114	MKLP	26	QKHGKPCDC	CSGG	26	KWKCRVTPKI	18	MHSSCYTEYQCRRIKNTYFTATL	21	QSPC-RGSINQVPC	WSDATAPIHIS	SDGGG	PL	TRKRV	VQKRL	190	RPKR	
avian	RevA	MDCL	35	QQLWEI	PCDCSGG	39	QWECVFKPKI	21	MHSTCYEKTQECTLLGKTYFTAIL	16	QASCTG-TVGKVP	CWSDVAPVYVSDGGG	PT	MIRES	VRELR	192	RHKR		
mouse	MMERV-β6	MLII	25	RAVYCN	PCDCSGG	34	SWIFVQKPN1	17	MNAICYTEYTDICIKDKAFFL	17	GLFSQKYP	YPIQLAFWPIAPMYSDG	G	PIQV	DLK	VQEWV	178	WHKR	
possum	TvERV(D)	MIST	23	HALFGAPDC	CKGG	34	SYQCVRVKI	20	LHSLCYTSTQCTGKSGTYLTSRQ	24	QASCDKINIG	KNVCSLHAPIHVS	SDGGG	PT	QIREM	EVKER	189	RPKR	
bat	DrERV	MFSQ	27	DALFGP	PCDCGG	36	RWVCVPKPRG	29	HSACCYKYVYTC	CSKGGQTYFTARI	28	PKCPG-EITG	KPACWKLDSPLG-	SDGGG	PQ	KRTER	VRIQI	179	RHKR

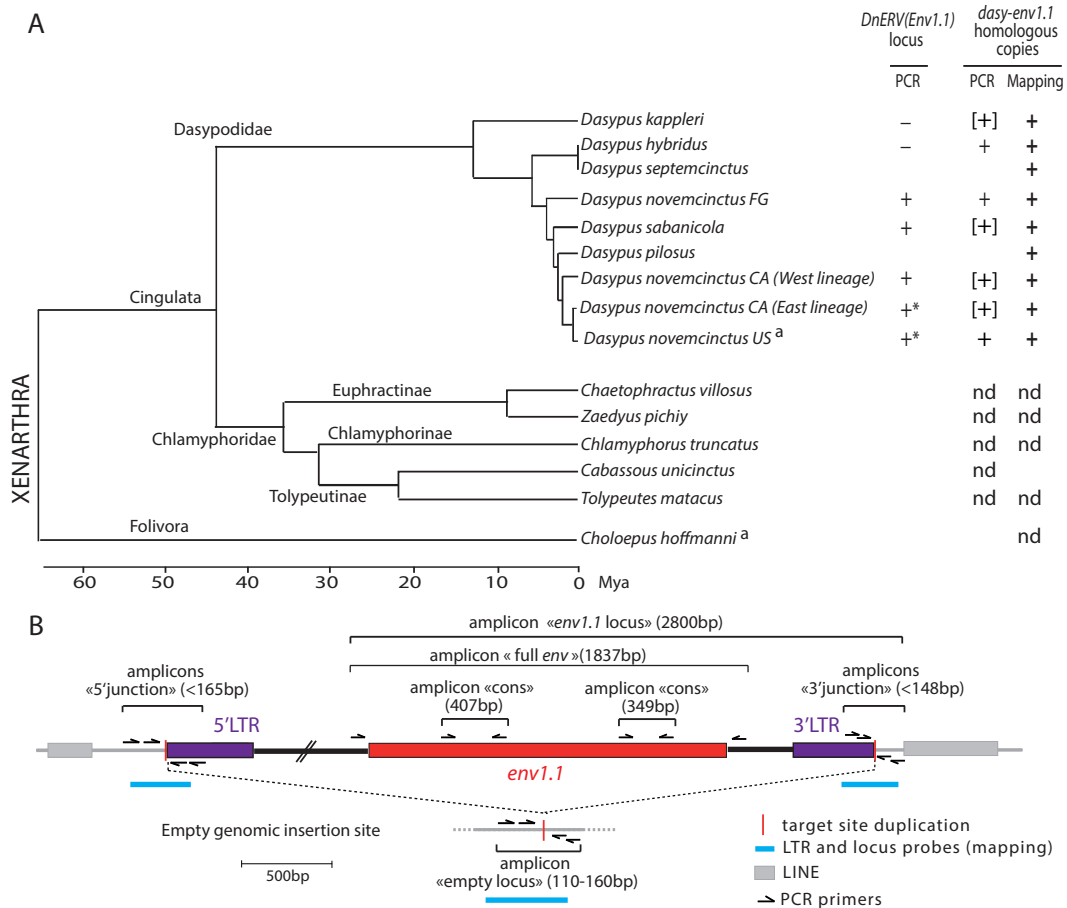
**FIG 7** Conservation of ASCT2-binding domains in the SU subunit of Dasy-Env1.1 and other retroviral Env proteins. Shown is a sequence alignment of SU protein subunits, from the methionine to the putative furin cleavage site. Virus names are indicated on the left, along with their host range. Variable regions are indicated, along with the number of omitted residues; dashes correspond to deletions. The conserved motif shown to be essential for the syncytin-1/hASCT2 interaction (SDGGGX2DXR) is highlighted in red, and the three cysteine-containing motifs (PCXC, CYX5C, and CX7-9CW) and the amino acids that are highly conserved among retroviral members of the type D interference group are highlighted in yellow (see reference 48). DrERV, coding sequence reconstituted from the DrERV\_216 noncoding Env sequence (GenBank accession number [KP175581](#)).

tion from Central America (Mexico, El Salvador, and Belize), 12 samples were heterozygous for the provirus, whereas 6 were negative, among which were the 2 individuals belonging to the “West” mitochondrial lineage. Insertional polymorphisms of the provirus were also detected in the *D. sabanicola* and *D. novemcinctus* FG populations (the latter possibly represents a distinct species [31]), with 3 and 2 individuals having one copy of the provirus, respectively, and 2 individuals of both species being negative. For the two

most basal species of the genus *Dasyus* (*D. kappleri* and *D. hybridus*), the empty-locus primer pairs amplified the expected amplicon in the 3 individuals tested, whereas no 5′ or 3′ proviral junction could be amplified in any of them. Although this seems to indicate the absence of the provirus in these species, more individuals need to be tested, as it remains possible that DnERV(Env1.1) is present in only a small fraction of individuals of these species. Mapping of Illumina reads (see below) to the



**FIG 8** DnERV(Env1.1) as well as a series of vertebrate endogenous and infectious retroviruses are chimeric retroviruses with a betaretrovirus *pol* gene and a gammaretrovirus *env* gene. The phylogenetic trees were determined as described in reference 11, by using the catalytic domain sequences of the RT region of the indicated viruses (left) or the TM region of the corresponding *env* genes (right) by the maximum likelihood method. The horizontal branch length and bar indicate the percentages of amino acid substitutions. Percent bootstrap values (>50%) obtained from 1,000 replicates are indicated at the nodes. Blue and red indicate typical gamma- and betaretroviruses, respectively, as determined by *pol* phylogenetic relationships, and orange indicates beta/gamma-type chimeric retroviruses. The following sequences were used: MoMLV (GenBank accession number [NC\\_001501.1](#)), FeLV (accession number [NC\\_001940.1](#)), PERV-A (accession numbers [AAM29192.1](#) and [AEF12608.1](#)), RD114 (accession number [NC\\_009889.1](#)), BaEV (accession number [D10032.1](#)), REV-A (accession number [DQ387450.1](#)), syncytin-1 (accession numbers [AAB66528.1](#) and [AF208161](#)), TvERV(D) (accession numbers [AF224725](#) and [AF284693](#)), SMRV (accession numbers [NC\\_001514.1](#) and [M23385.1](#)), DrERV (accession numbers [KP175580](#) and [KP175581](#)), syncytin-Ory1-ERV (accession numbers [XP\\_008268050](#) and [ACZ58381](#)), SRV1 (accession number [M11841.1](#)), SRV4 (accession number [NC\\_014474.1](#)), MPMV (accession number [NC\\_001550.1](#)), MMERV-β6 (accession number [NT\\_039167](#)), HERV-K (accession number [AY037928.1](#)), MMTV (accession number [NC\\_001503.1](#)), enJSRV (accession numbers [AAD45226.1](#) and [AAF22166.1](#)), and HIV-1 (accession number [K02013](#)).



**FIG 9** Identification and conservation of *dasy-env1.1* and its related elements during xenarthran radiation. (A, left) Phylogenetic tree of Xenarthra (31) with the names of the species tested together with their corresponding families and subfamilies. The horizontal branch length is proportional to time. FG, French Guiana; CA, Central America. <sup>a</sup>, species whose genome is available in genomic databases. (Right) The presence (+) or absence (-) of the DnERV(Env1.1) orthologous locus or of *dasy-env1.1* homologous copies determined by using either PCR assays ("PCR") or mapping on genomic Illumina reads ("Mapping") is indicated for each species tested (the individual status of each species is provided in Table 1). Brackets indicate that only a partial sequence could be retrieved. Asterisks indicate that a full-length *dasy-env1.1* ORF was detected at the orthologous locus in individuals of this species (Table 1). nd, not detected. (B) Strategy used to detect the DnERV(Env1.1) provirus at its orthologous locus as well as provirus-related sequences. The presence or absence of a proviral locus was determined by using three PCRs (with a limited elongation time): the "5' junction" and "3' junction" primer pairs amplify sequences when DnERV is present at the locus. If the provirus is not integrated at the locus, the "empty locus" primer pair, but not the two other primer combinations, yields a PCR product (corresponding to the joined flanking sequences). If all 3 primer pairs yield a product, the provirus insertion is heterozygous. If no primer pairs amplify a sequence, the quality of the DNA may be too poor. The presence of *dasy-env1.1* at the orthologous locus was determined by using the "env1.1 locus" primer pair. The presence or absence of DnERV(Env1.1)-related sequences was determined by using three PCRs with two "cons" primer pairs and the "full env" primer pair, which amplify partial and full-length *env* fragments, respectively. The schematized amplicons are not drawn to scale. Finally, the positions of the probes that were used for the mapping of the shotgun Illumina genomic reads are indicated.

empty proviral locus further confirmed the presence of alleles with empty preintegration loci in the genomes of individuals belonging to the *D. novemcinctus* Central American and French Guiana populations and to *D. sabanicola* populations (Table 1). The full-length *dasy-env1.1* gene was then tentatively amplified at the orthologous locus from the genomic DNAs of all *Dasyus* species using a locus-specific pair of primers (designated "env1.1 locus") (Fig. 9B). In 12 *D. novemcinctus* samples from El Salvador, Belize, and the United States, a PCR product of the expected size was obtained, confirming the presence of the *dasy-env1.1* ortholog in various members of the *D. novemcinctus* lineage. Sequencing of the PCR products revealed the presence of a full-length *dasy-env1.1* ORF (586 to 592 aa long) (Table 1 and Fig. 9). Altogether, these data suggest that the orthologous DnERV(Env1.1) provirus is present mostly as a heterozygous copy in the genus *Dasyus*

(only one U.S. individual harbored a homozygous proviral insertion) and that its insertion is polymorphic, with 14 individuals being devoid of the proviral insertion among the U.S., Central American, and South American *Dasyus* populations.

Individuals negative for the DnERV(Env1.1) provirus were then investigated for the presence of other DnERV(Env1.1)-related sequences. Low-stringency PCR assays to amplify homologous full-length *dasy-env1.1* sequences from these DNAs using primers within the provirus and bracketing the *dasy-env1.1* gene (designated "full env") (Fig. 9B) yielded a single amplicon of the expected size in *D. novemcinctus* from French Guiana, which represents a distinct mitochondrial lineage (31), and in *D. hybridus*, in addition to the *D. novemcinctus* U.S. lineage, as expected (Fig. 9). Sequencing of the PCR products after cloning of the amplicons in the pGEM-T vector revealed several distinct full-length *dasy-*

*env1.1*-related sequences with >95% identity but with frameshifts and/or stop codons, consistent with the presence of noncoding sequences in these species. Finally, two primer pairs designed internally to the *env* ORF in the most conserved regions of retroviral *env* genes (namely, the furin cleavage site and the C-X<sub>5,6,7</sub>-C motif) (designated “cons” primers) (Fig. 9B) successfully amplified fragments homologous to *dasy-env1.1* within the DNAs of the *D. novemcinctus* CA West and East lineages, *D. sabanicola*, and the most basal species of *Dasyapus*, *D. kappleri*, but not within the DNAs of the species of the more distantly related subfamilies belonging to the Chlamyphoridae family (from the same order, Cingulata, as Dasypodidae), i.e., Euphractinae, Chlamyphorinae, and Tolypeutinae (whereas, accordingly, the empty locus could be amplified in these species [not shown]) (Fig. 9A and Table 1). This indicates that the DnERV-1 *env* gene family is present only in species belonging to the Dasypodidae family. This was confirmed by mapping of low-coverage shotgun Illumina genomic reads for all Xenarthra species (5 million to 30 million Illumina reads [31]). Sequences homologous to DnERV(Env1.1) LTRs or *env* were found in several Dasypodidae species (*D. kappleri*, *D. hybridus*, *D. septemcinctus*, *D. novemcinctus* FG, *D. sabanicola*, *D. pilosus*, and *D. novemcinctus*) but not in any other subfamily representatives (Fig. 9 and Table 1). Finally, no sequence with >50% similarity to *dasy-env1.1* could be found within the genome of the two-toed sloth (*Choloepus hoffmanni*) (Ensembl databases) by using an *in silico* BLAST search (Fig. 9A).

Altogether, these data suggest that (i) elements of the DnERV-1 family originally inserted into the genome of the *Dasy-* ancestral lineage between 12 Mya and 45 Mya and (ii) the DnERV(Env1.1) provirus inserted into the identified locus, at least before the divergence between the *D. novemcinctus* FG lineage and other armadillos, estimated at around 3.7 Mya (31).

## DISCUSSION

In this study, using *in silico* screening, we provide comprehensive characterization and quantification of the retroviral Env protein-encoding genes residing in the sequenced nine-banded armadillo genome. A BLAST search identified 162 putative retroviral *env* genes. This number far exceeds those found in the genomes of other mammalian species for which we performed a strictly similar screening (i.e., 9 in *Echinops telfairi* [40] and 18 in *Bos taurus* [51]). This is consistent with data from a previous study by Hayward et al. (10), who screened vertebrate genomes for the presence of high-quality ERVs containing conserved *gag* and *pol* motifs and reported a higher number of ERVs in the genome of the nine-banded armadillo than in the genomes of many other mammals (10). A large fraction, i.e., 87% (140/162), of these *env* genes could be assigned to class I ERVs (gammaretroviruses), and only 13% (22/162) were assigned to class II ERVs (betaretroviruses), with no other retroviral genera being represented. It is noteworthy that the relative fraction of betaretrovirus-like *env* sequences is small compared to that of *gag-pol* sequences of a betaretroviral origin reported by Hayward et al., i.e., 72% (10). This can be accounted for by a significant fraction of the armadillo genome being occupied by recombinant beta/gamma-type ERVs and/or by *env*-less betaretrovirus-like ERV families, which have been reported to expand in certain mammalian species (52) and whose quantitative importance in armadillo remains to be addressed. Finally, the abundance of *env* genes in the armadillo genome results largely from a small number of significant bursts of closely related ERVs

(e.g., *Dasy-Env5* to *-Env14*). These intense bursts, which have been observed for other mammalian species, e.g., mouse (53), suggest waves of expansion by reinfection events, which might be correlated with a reduced capacity of host control.

Among the 162 candidate *env* genes, we identified one, *dasy-env1.1*, that is phylogenetically related to two previously characterized *syncytin* genes and that possesses all the putative canonical domains for a functional *env* gene. We demonstrated that it displays fusogenic properties by both *ex vivo* pseudotyping and cell-cell fusion assays. Moreover, by using cloning and coexpression experiments, *Dasy-Env1.1* was found to use the ASCT2 transporter as the receptor mediating cell fusion, as also demonstrated for the rabbit *syncytin-Ory1* and human *syncytin-1* genes; for ERVs/retroviruses of the RD114 retrovirus interference group in diverse vertebrate species; and, as we now propose, for three previously described ERVs from bat, possum, and mouse (see Results). The discovery of this *env* gene, which is not a canonical *syncytin* gene (see below), has several interesting implications.

First, the present study points out that the ASCT2 transporter is a “successful” receptor, whose function has been repeatedly subverted by independent retroviruses to gain entry into host cells of various species and become endogenized. We showed that the *Dasy-Env1.1* protein is able to functionally interact with both the nine-banded armadillo and the human ASCT2 transporters, as does human *syncytin-1*. This strong conservation of ASCT2 receptor recognition throughout mammalian evolution (see also references 54 and 55) probably explains the susceptibility of various vertebrate species to ASCT2-mediated retroviral endogenization, a feature that is now extended to the xenarthran clade of placental mammals. Moreover, as similarly observed for the human, murine, bovine, monkey, and canine ASCT2 genes (54, 56–58), *dasyASCT2* is widely expressed in several armadillo tissues, with its transcripts being detected in spleen, liver, kidney, muscle, and placenta (Fig. 2B). In humans, consistent with its specific role in retroviral endogenization events, the ASCT2 transporter was recently found to be present in gametes (spermatozoa and oocytes) (59), and we detected ASCT2 transcripts in single-cell transcriptome sequencing (RNA-Seq) data generated from human preimplantation embryos (60), from mature oocytes to late blastocysts (not shown). All these factors likely contribute to the repeated ASCT2 receptor use in retroviral endogenization events.

Second, *dasy-env1.1*, which is phylogenetically related to gammaretroviral *env* (e.g., REV-A, BaEV, and MoMLV), belongs to a provirus that has typical betaretroviral characteristics, suggesting recombination events. Interestingly, we further showed that *Dasy-Env1.1* displays phylogenetic proximity to a large series of gamma-type Env proteins, which have also been acquired by recombinant betaretroviruses in various vertebrate species and which all share the SDGGGX2DX2R motif involved in *syncytin-1*/hASCT2 interactions. This is the case for the simian retroviruses SMRV, SRV, and MPMV (48) as well as for rabbit *syncytin-Ory1*-ERV, bat DrERV, possum TvERV(D), and murine MMERV-β6, as revealed in this study. While other recombinants between beta- and gamma-type retroviruses have been described (e.g., *Python molurus* endogenous retrovirus [PyERV] [61], intracisternal A-particle Env-encoding proviral elements in shrew and guinea pig [52], and marsupial Opo-Env3-ERV [28]), the acquisition of a gammaretroviral *env* gene that specifically binds and infects cells expressing the ASCT2 receptor therefore turns out to be a frequent event. We propose that ASCT2 usage is one of the driving

forces for the recombination events leading to the emergence of DnERV and of other chimeric viruses in a panel of vertebrate species. Given the strong conservation and ubiquitous expression of this transporter, viruses using this receptor most likely have a wide host range, with the potential to jump between species and to infect multiple tissues, which might contribute to their survival and transmission. As a corollary, such infectious retroviruses or reemerging ERVs circulating in the wild, e.g., in the nine-banded armadillo, or even domestic animal species can be considered potentially dangerous zoonotic pathogens for human populations. The recent discovery in humans of a secreted HERV Env-derived protein, suppressyn, which is able to compete for binding to the ASCT2 receptor and which potentially acts as a restriction factor via receptor interference (62), suggests mechanisms that evolved to control this threat.

Third, the *dasy-env1.1* gene belongs to an endogenous retrovirus, the DnERV(Env1.1) provirus, whose endogenization within the *Dasyopus* population has not reached fixation. Based on the presence or absence of DnERV(Env1.1) at orthologous loci and estimated dates of speciation, DnERV(Env1.1) integrated into the *Dasyopus* lineage prior to the divergence of *D. novemcinctus* FG and other *Dasyopus* species around 3.7 Mya (31). This is congruent with our age estimate using LTR sequence divergence, suggesting an insertion of the DnERV(Env1.1) provirus approximately 5.5 Mya. Other family members integrated before diversification of the genus *Dasyopus* >12 Mya. However, testing of several individuals of the same species showed that the DnERV(Env1.1) integration is not present in all members of the species *D. novemcinctus*, *D. sabanicola*, and *D. novemcinctus* FG (insertion frequency ranging from 50 to 86%), indicating insertional polymorphism among members of the *Dasyopus* genus. Moreover, DnERV(Env1.1) is present mostly as a haploid copy, consistent with the fact that it has not yet reached fixation. Indeed, only 1 individual, belonging to the *D. novemcinctus* U.S. lineage, was found to be homozygous for the insertion, among the 24 positive individuals tested. Of note, orthologous *dasy-env1.1* copies with full coding capacity were detected in several *D. novemcinctus* individuals from the United States and Central America.

The long-term persistence of the unfixed DnERV(Env1.1) provirus suggests that its insertion has no deleterious phenotypic consequence, i.e., that it is selectively neutral, although the hypothesis that selection would favor individuals that express the *dasy-env1.1* fusogenic gene cannot be ruled out (see below). Parameters that drive ERV loci to reach high allele frequencies and eventually become fixed within a host population are still poorly understood (63, 64). Due to our small sample size, it is not possible to determine whether the number of individuals that are homozygous for the DnERV(Env1.1) insertion is statistically higher among the U.S. population (which contains the only individual that is homozygous for the insertion) than in the non-U.S. ones. Additional studies are required to determine whether the original founder effect that occurred during armadillo colonization of the United States, associated with possible inbreeding events and illustrated by the genetic homogeneity among U.S. individuals (65), would finally favor the fixation of the provirus in this specific population. A few other well-documented examples of insertional polymorphism of ERVs have been reported for koalas (14), cats (17), pigs (66), mice (67), and cervids (16, 18). Although commonly associated with evolutionarily younger ERV integrants, examples of "older" ERVs, as observed in armadillos, are also found in hu-

mans, where the full-length HERV-K115 provirus, although having been integrated more than 1.1 Mya, displays an average insertion frequency of 28%, with no individuals being found to be homozygous for this insertion (68). In the future, a thorough analysis of the prevalence and distribution of *dasy-env1.1*-related copies among a wide range of *Dasyopus* individuals, and in relation to population dynamics, genetic diversity, and range expansion, would provide new insights into both ERV evolutionary rates and the fine structure and history of contemporary *Dasyopus* populations.

Finally, one of the outcomes of this study is that no bona fide *syncytin* genes emerged from our screening of the nine-banded armadillo genome. Syncytins are defined by three major characteristic properties: placenta-specific expression, fusogenic properties, and conservation in evolution for an extended period of time (usually >12 million years) (27). In fact, among the 162 putative *env* genes that were identified, no candidate *env* gene, including the *dasy-env1* family, was found to be preferentially expressed in the placenta, although we could not formally reject candidate *syncytin* genes based on this criterion, as only multicopy transcript levels could be assessed by qRT-PCR analysis. The *dasy-env1.1* gene, which was selected based upon its phylogenetic relationships with *syncytin* genes and its fusogenic properties (see below), inserted into the *Dasyopus* lineage relatively recently (ca. 5 Mya) and did not yet reach fixation, suggesting that it did not evolve under strong selective pressure. Finally, all other candidate *env* genes, apart from *dasy-env1.1*, that have been tentatively tested for fusogenicity, i.e., prototype sequences of the *dasy-env2*, *-env3*, and *-env4* families, were found to be negative (data not shown). Although we cannot exclude the absence of a *syncytin* gene in this species, one hypothesis is that the incomplete coverage of the *D. novemcinctus* genome sequence assembly might preclude the identification of such a gene, reported in most cases to be present as a single copy (27). However, the identified *dasy-env1.1* gene displays some of the properties of a *syncytin* gene. It exhibits cell-cell fusogenic properties and binds the same cellular receptor, ASCT2, as two other *syncytin* genes (see above). Moreover, transcripts homologous to *dasy-env1.1* can be detected by *in situ* hybridization in the placenta, specifically at the level of cytotrophoblasts, which normally fuse to form the syncytiotrophoblast (data not shown). Although DnERV(Env1.1) integrated into the *Dasyopus* genus relatively recently, one could hypothesize that its *env* gene, *dasy-env1.1*, or one of its related copies, currently contributes to placental syncytium formation and may turn out to be a future *syncytin* gene. The possibility that *dasy-env1.1* could act as a restriction factor, by blocking receptors against related infectious retroviruses (20), might also be considered. In this respect, it should be noted that orthologous *dasy-env1.1* genes have conserved their full coding capacity among *D. novemcinctus* individuals from the United States and Central America.

Some molecular events have been shown to be associated with the acquisition of a *syncytin* function. Recruitment of upstream regulatory elements in the vicinity of the integration site (69) or acquisition of a new non-LTR promoter (70) probably conferred placenta-specific expression to some *syncytin* genes. The *dasy-env1.1* ORF displays a 4-bp C-terminal duplication, which is not found in any other members of this family whose sequences are available in genomic databases (data not shown) and which results in a 5-aa truncation of the cytoplasmic C-terminal tail due to premature translation termination. This could have enhanced the



fusogenic capacity of Dasy-Env1.1, as observed for several gamma- and beta-type Env proteins, which can be rendered fusogenic via C-terminal truncation (71). In this respect, *dasy-env1.1*, thanks to future modifications and provided that it confers a selective advantage, could progressively replace an existing, but as-yet-unidentified, *syncytin* gene in *Dasybus* and contribute to the turnover of these genes throughout evolution (27, 72, 73).

## ACKNOWLEDGMENTS

We especially thank F. Knight (University of the Ozarks, Clarksville, AR, USA) as well as F. Catzeflis (Institut des Sciences de l'Évolution, Montpellier, France), W. J. Loughry, C. McDonough (Valdosta State University, Valdosta, GA, USA), R. Truman (National Hansen's Disease Program Laboratory, Baton Rouge, LA, USA), F. A. Jimenez (Southern Illinois University, Carbondale, IL, USA), B. McAllister (University of Iowa, Iowa City, IA, USA), J. Pérez (Parque Zoológico Nacional de El Salvador), the Museum of Southwestern Biology (Albuquerque, NM, USA), L. Sigler, A. Noss, M.-C. Arteaga, C. Silva, M. Superina, and D. Devillier for armadillo capture and tissue collection; R. Potier (ZooParc de Beauval) for blood samples; O. Bawa for her contribution to the histological analyses; and D. Mager and G. Baillie for providing us published sequences. We thank Christian Lavielle for comments and critical reading of the manuscript and three anonymous reviewers for helpful suggestions.

This work was supported by the CNRS and by grants from the Ligue Nationale contre le Cancer (T.H.; Equipe Labelisée) and the Agence Nationale de la Recherche (ANR Retro-Placenta and CEBA ANR-10-LABX-25-01).

## FUNDING INFORMATION

This work, including the efforts of Sebastien Malicorne, Cécile Vernochet, Guillaume Cornelis, Odile Heidmann, Thierry Heidmann, and Anne Dupressoir, was funded by Agence Nationale de la Recherche (ANR) (RETRO-PLACENTA).

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## REFERENCES

- Zhuo X, Feschotte C. 2015. Cross-species transmission and differential fate of an endogenous retrovirus in three mammal lineages. *PLoS Pathog* 11:e1005279. <http://dx.doi.org/10.1371/journal.ppat.1005279>.
- Jern P, Coffin JM. 2008. Effects of retroviruses on host genome function. *Annu Rev Genet* 42:709–732. <http://dx.doi.org/10.1146/annurev.genet.42.110807.091501>.
- Stoye JP. 2012. Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nat Rev Microbiol* 10:395–406. <http://dx.doi.org/10.1038/nrmicro2783>.
- Mager DL, Stoye JP. 2016. Mammalian endogenous retroviruses. *Microbiol Spectr* 3(1):MDNA3-0009-2014. <http://dx.doi.org/10.1128/microbiolspec.MDNA3-0009-2014>.
- Overbaugh J, Miller AD, Eiden MV. 2001. Receptors and entry cofactors for retroviruses include single and multiple transmembrane-spanning proteins as well as newly described glycoposphatidylinositol-anchored and secreted proteins. *Microbiol Mol Biol Rev* 65:371–389. <http://dx.doi.org/10.1128/MMBR.65.3.371-389.2001>.
- Manel N, Kim FJ, Kinet S, Taylor N, Sitbon M, Battini JL. 2003. The ubiquitous glucose transporter GLUT-1 is a receptor for HTLV. *Cell* 115:449–459. [http://dx.doi.org/10.1016/S0092-8674\(03\)00881-X](http://dx.doi.org/10.1016/S0092-8674(03)00881-X).
- Sommerfelt MA, Weiss RA. 1990. Receptor interference groups of 20 retroviruses plating on human cells. *Virology* 176:58–69. [http://dx.doi.org/10.1016/0042-6822\(90\)90230-O](http://dx.doi.org/10.1016/0042-6822(90)90230-O).
- Taylor CS, Nouri A, Zhao Y, Takeuchi Y, Kabat D. 1999. A sodium-dependent neutral-amino-acid transporter mediates infections of feline and baboon endogenous retroviruses and simian type D retroviruses. *J Virol* 73:4470–4474.
- Gifford R, Tristem M. 2003. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* 26:291–315. <http://dx.doi.org/10.1023/A:1024455415443>.
- Hayward A, Cornwallis C, Jern P. 2015. Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. *Proc Natl Acad Sci U S A* 112:464–469. <http://dx.doi.org/10.1073/pnas.1414980112>.
- Benit L, Dessen P, Heidmann T. 2001. Identification, phylogeny, and evolution of retroviral elements based on their envelope genes. *J Virol* 75:11709–11719. <http://dx.doi.org/10.1128/JVI.75.23.11709-11719.2001>.
- Vargiu L, Rodriguez-Tome P, Sperber GO, Cadeddu M, Grandi N, Blikstad V, Tramontano E, Blomberg J. 2016. Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology* 13:7. <http://dx.doi.org/10.1186/s12977-015-0232-y>.
- Henzy JE, Johnson WE. 2013. Pushing the endogenous envelope. *Philos Trans R Soc Lond B Biol Sci* 368:20120506. <http://dx.doi.org/10.1098/rstb.2012.0506>.
- Tarlinton RE, Meers J, Young PR. 2006. Retroviral invasion of the koala genome. *Nature* 442:79–81. <http://dx.doi.org/10.1038/nature04841>.
- Arnaud F, Caporale M, Varela M, Biek R, Chessa B, Alberti A, Golder M, Mura M, Zhang Y, Yu L, Pereira F, DeMartini JC, Leymaster K, Spencer TE, Palmarini M. 2007. A paradigm for virus-host coevolution: sequential counter-adaptations between endogenous and exogenous retroviruses. *PLoS Pathog* 3:e170. <http://dx.doi.org/10.1371/journal.ppat.0030170>.
- Elleder D, Kim O, Padhi A, Bankert JG, Simeonov I, Schuster SC, Wittekindt NE, Motameny S, Poss M. 2012. Polymorphic integrations of an endogenous gammaretrovirus in the mule deer genome. *J Virol* 86:2787–2796. <http://dx.doi.org/10.1128/JVI.06859-11>.
- Anai Y, Ochi H, Watanabe S, Nakagawa S, Kawamura M, Gojobori T, Nishigaki K. 2012. Infectious endogenous retroviruses in cats and emergence of recombinant viruses. *J Virol* 86:8634–8644. <http://dx.doi.org/10.1128/JVI.00280-12>.
- Kamath PL, Elleder D, Bao L, Cross PC, Powell JH, Poss M. 2014. The population history of endogenous retroviruses in mule deer (*Odocoileus hemionus*). *J Hered* 105:173–187. <http://dx.doi.org/10.1093/jhered/est088>.
- Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I, Kao RR, Pemberton JM, Beraldi D, Stear MJ, Alberti A, Pittau M, Iannuzzi L, Banabazi MH, Kazwala RR, Zhang YP, Arranz JJ, Ali BA, Wang Z, Uzun M, Dione MM, Olsaker I, Holm LE, Saarma U, Ahmad S, Marzanov N, Eythorsdottir E, Holland MJ, Ajmone-Marsan P, Bruford MW, Kantanen J, Spencer TE, Palmarini M. 2009. Revealing the history of sheep domestication using retrovirus integrations. *Science* 324:532–536. <http://dx.doi.org/10.1126/science.1170587>.
- Aswad A, Katourakis A. 2012. Paleovirology and virally derived immunity. *Trends Ecol Evol* 27:627–636. <http://dx.doi.org/10.1016/j.tree.2012.07.007>.
- Mi S, Lee X, Li X, Veldman G, Finnerty H, Racie L, LaVallie E, Tang X, Edouard P, Howes S, Keith JJ, McCoy J. 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403:785–789. <http://dx.doi.org/10.1038/35001608>.
- Blond JL, Lavillette D, Cheynet V, Bouton O, Oriol G, Chapel-Fernandes S, Mandrand B, Mallet F, Cosset FL. 2000. An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. *J Virol* 74:3321–3329. <http://dx.doi.org/10.1128/JVI.74.7.3321-3329.2000>.
- Blaise S, de Parseval N, Benit L, Heidmann T. 2003. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc Natl Acad Sci U S A* 100:13013–13018. <http://dx.doi.org/10.1073/pnas.2132646100>.
- Dupressoir A, Marceau G, Vernochet C, Benit L, Kanellopoulos C, Sapin V, Heidmann T. 2005. Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. *Proc Natl Acad Sci U S A* 102:725–730. <http://dx.doi.org/10.1073/pnas.0406509102>.
- Dupressoir A, Vernochet C, Bawa O, Harper F, Pierron G, Opolon P, Heidmann T. 2009. Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc Natl Acad Sci U S A* 106:12127–12132. <http://dx.doi.org/10.1073/pnas.0902925106>.
- Dupressoir A, Vernochet C, Harper F, Guégan J, Dessen P, Pierron G, Heidmann T. 2011. A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast. *Proc Natl Acad Sci U S A* 108:1164–1173. <http://dx.doi.org/10.1073/pnas.1012185108>.
- Lavielle C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Verno-

- chet C, Heidmann T. 2013. Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. *Philos Trans R Soc Lond B Biol Sci* 368:20120507. <http://dx.doi.org/10.1098/rstb.2012.0507>.
28. Cornelis G, Vernochet C, Carradec Q, Souquere S, Mulot B, Catzeflis F, Nilsson MA, Menzies BR, Renfree MB, Pierron G, Zeller U, Heidmann O, Dupressoir A, Heidmann T. 2015. Retroviral envelope gene captures and syncytin exaptation for placentation in marsupials. *Proc Natl Acad Sci U S A* 112:E487–E496. <http://dx.doi.org/10.1073/pnas.1417000112>.
  29. Heidmann O, Vernochet C, Dupressoir A, Heidmann T. 2009. Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new "syncytin" in a third order of mammals. *Retrovirology* 6:107. <http://dx.doi.org/10.1186/1742-4690-6-107>.
  30. Katzourakis A. 2013. Paleovirology: inferring viral evolution from host genome sequence data. *Philos Trans R Soc Lond B Biol Sci* 368:20120493. <http://dx.doi.org/10.1098/rstb.2012.0493>.
  31. Gibb GC, Condamine FL, Kuch M, Enk J, Moraes-Barros N, Superina M, Poinar HN, Delsuc F. 2016. Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans. *Mol Biol Evol* 33:621–642. <http://dx.doi.org/10.1093/molbev/msv250>.
  32. Taulman JF, Robbins LW. 2014. Range expansion and distributional limits of the nine-banded armadillo in the United States: an update of Taulman & Robbins (1996). *J Biogeogr* 41:1626–1630. <http://dx.doi.org/10.1111/jbi.12319>.
  33. Hashemolhosseini S, Wegner M. 2004. Impacts of a new transcription factor family: mammalian GCM proteins in health and disease. *J Cell Biol* 166:765–768. <http://dx.doi.org/10.1083/jcb.200406097>.
  34. Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27:221–224. <http://dx.doi.org/10.1093/molbev/msp259>.
  35. Stamatakis A. 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690. <http://dx.doi.org/10.1093/bioinformatics/btl446>.
  36. de Parseval N, Lazar V, B nit L, Casella J, 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* 77:10414–10422. <http://dx.doi.org/10.1128/JVI.77.19.10414-10422.2003>.
  37. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. <http://dx.doi.org/10.1093/bioinformatics/bts199>.
  38. Esnault C, Priet S, Ribet D, Vernochet C, Bruls T, Lavialle C, Weissenbach J, Heidmann T. 2008. A placenta-specific receptor for the fusogenic, endogenous retrovirus-derived, human syncytin-2. *Proc Natl Acad Sci U S A* 105:17532–17537. <http://dx.doi.org/10.1073/pnas.0807413105>.
  39. Blaise S, de Parseval N, Heidmann T. 2005. Functional characterization of two newly identified human endogenous retrovirus coding envelope genes. *Retrovirology* 2:19. <http://dx.doi.org/10.1186/1742-4690-2-19>.
  40. Cornelis G, Vernochet C, Malicorne S, Souquere S, Tzika AC, Goodman SM, Catzeflis F, Robinson TJ, Milinkovitch MC, Pierron G, Heidmann O, Dupressoir A, Heidmann T. 2014. Retroviral envelope syncytin capture in an ancestrally diverged mammalian clade for placentation in the primitive Afrotherian tenrecs. *Proc Natl Acad Sci U S A* 111:E4332–E4341. <http://dx.doi.org/10.1073/pnas.1412268111>.
  41. Hizi A, Herzog E. 2015. dUTPase: the frequently overlooked enzyme encoded by many retroviruses. *Retrovirology* 12:70. <http://dx.doi.org/10.1186/s12977-015-0198-9>.
  42. Chiu IM, Skuntz SF. 1986. Nucleotide sequence analysis of squirrel monkey retrovirus reveals a novel primer-binding site for tRNA<sup>Lys1,2</sup>. *J Virol* 58:983–987.
  43. Katzourakis A, Gifford R, Tristem M, Gilbert M, Pybus O. 2009. Macroevolution of complex retroviruses. *Science* 325:1512. <http://dx.doi.org/10.1126/science.1174149>.
  44. Henzy JE, Coffin JM. 2013. Betaretroviral envelope subunits are noncovalently associated and restricted to the mammalian class. *J Virol* 87:1937–1946. <http://dx.doi.org/10.1128/JVI.01442-12>.
  45. Escalera-Zamudio M, Mendoza ML, Heeger F, Loza-Rubio E, Rojas-Anaya E, Mendez-Ojeda ML, Taboada B, Mazzoni CJ, Arias CF, Greenwood AD. 2015. A novel endogenous betaretrovirus in the common vampire bat (*Desmodus rotundus*) suggests multiple independent infection and cross-species transmission events. *J Virol* 89:5180–5184. <http://dx.doi.org/10.1128/JVI.03452-14>.
  46. Baillie GJ, Wilkins RJ. 2001. Endogenous type D retrovirus in a marsupial, the common brushtail possum (*Trichosurus vulpecula*). *J Virol* 75:2499–2507. <http://dx.doi.org/10.1128/JVI.75.5.2499-2507.2001>.
  47. Baillie GJ, van de Lagemaat LN, Baust C, Mager DL. 2004. Multiple groups of endogenous betaretroviruses in mice, rats, and other mammals. *J Virol* 78:5784–5798. <http://dx.doi.org/10.1128/JVI.78.11.5784-5798.2004>.
  48. Cheynet V, Oriol G, Mallet F. 2006. Identification of the hASCT2-binding domain of the Env ERVWE1/syncytin-1 fusogenic glycoprotein. *Retrovirology* 3:41. <http://dx.doi.org/10.1186/1742-4690-3-41>.
  49. Sonigo P, Barker C, Hunter E, Wain-Hobson S. 1986. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. *Cell* 45:375–385. [http://dx.doi.org/10.1016/0092-8674\(86\)90323-5](http://dx.doi.org/10.1016/0092-8674(86)90323-5).
  50. Arteaga M, Pi nero D, Eguiarte L, Gasca J, Medellin R. 2012. Genetic structure and diversity of the nine-banded armadillo in Mexico. *J Mammal* 93:547–559. <http://dx.doi.org/10.1644/11-MAMM-A-211.1>.
  51. Cornelis G, Heidmann O, Degrelle S, Vernochet C, Lavialle C, Letzelter C, Bernard-Stoecklin S, Hassanin A, Mulot B, Guillomot M, Hue I, Heidmann T, Dupressoir A. 2013. Captured retroviral envelope syncytin gene associated with the unique placental structure of higher ruminants. *Proc Natl Acad Sci U S A* 110:E828–E837. <http://dx.doi.org/10.1073/pnas.1215787110>.
  52. Magiorakis G, Gifford RJ, Katzourakis A, De Ranter J, Belshaw R. 2012. Env-less endogenous retroviruses are genomic superspreaders. *Proc Natl Acad Sci U S A* 109:7385–7390. <http://dx.doi.org/10.1073/pnas.1200913109>.
  53. Benit L, Lallemand JB, Casella JF, Philippe H, Heidmann T. 1999. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. *J Virol* 73:3301–3308.
  54. Yoshikawa R, Yasuda J, Kobayashi T, Miyazawa T. 2012. Canine ASCT1 and ASCT2 are functional receptors for RD-114 virus in dogs. *J Gen Virol* 93:603–607. <http://dx.doi.org/10.1099/vir.0.036228-0>.
  55. Miyaho RN, Nakagawa S, Hashimoto-Gotoh A, Nakaya Y, Shimode S, Sakaguchi S, Yoshikawa R, Takahashi MU, Miyazawa T. 2015. Susceptibility of domestic animals to a pseudotype virus bearing RD-114 virus envelope protein. *Gene* 567:189–195. <http://dx.doi.org/10.1016/j.gene.2015.04.079>.
  56. Green BJ, Lee CS, Rasko JE. 2004. Biodistribution of the RD114/mammalian type D retrovirus receptor, RDR. *J Gene Med* 6:249–259. <http://dx.doi.org/10.1002/jgm.517>.
  57. Ochiai H, Onda K, Ogihara K, Naya Y, Sugiyama H, Maruo T. 2012. cDNA sequence and tissue distribution of canine Na-dependent neutral amino acid transporter 2 (ASCT2). *J Vet Med Sci* 74:1505–1510. <http://dx.doi.org/10.1292/jvms.12-0171>.
  58. Shimode S, Nakaoka R, Shogen H, Miyazawa T. 2013. Characterization of feline ASCT1 and ASCT2 as RD-114 virus receptor. *J Gen Virol* 94:1608–1612. <http://dx.doi.org/10.1099/vir.0.052928-0>.
  59. Bjerregaard B, Lemmen JG, Petersen MR, Ostrup E, Iversen LH, Almstrup K, Larsson LI, Ziehe S. 2014. Syncytin-1 and its receptor is present in human gametes. *J Assist Reprod Genet* 31:533–539. <http://dx.doi.org/10.1007/s10815-014-0224-1>.
  60. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J, Huang J, Li M, Wu X, Wen L, Lao K, Qiao J, Tang F. 2013. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol* 20:1131–1139. <http://dx.doi.org/10.1038/nsmb.2660>.
  61. Huder JB, Boni J, Hatt JM, Soldati G, Lutz H, Schupbach J. 2002. Identification and characterization of two closely related unclassifiable endogenous retroviruses in pythons (*Python molurus* and *Python curtus*). *J Virol* 76:7607–7615. <http://dx.doi.org/10.1128/JVI.76.15.7607-7615.2002>.
  62. Sugimoto J, Sugimoto M, Bernstein H, Jinno Y, Schust D. 2013. A novel human endogenous retroviral protein inhibits cell-cell fusion. *Sci Rep* 3:1462. <http://dx.doi.org/10.1038/srep01462>.
  63. Katzourakis A, Pereira V, Tristem M. 2007. Effects of recombination rate on human endogenous retrovirus fixation and persistence. *J Virol* 81:10712–10717. <http://dx.doi.org/10.1128/JVI.00410-07>.
  64. Kanda RK, Tristem M, Coulson T. 2013. Exploring the effects of immunity and life history on the dynamics of an endogenous retrovirus. *Philos Trans R Soc Lond B Biol Sci* 368:20120505. <http://dx.doi.org/10.1098/rstb.2012.0505>.
  65. Huchon D, Delsuc F, Catzeflis FM, Douzery EJ. 1999. Armadillos

- exhibit less genetic polymorphism in North America than in South America: nuclear and mitochondrial data confirm a founder effect in *Dasyurus novemcinctus* (Xenarthra). *Mol Ecol* 8:1743–1748. <http://dx.doi.org/10.1046/j.1365-294x.1999.00768.x>.
66. Lavillette D, Kabat D. 2004. Porcine endogenous retroviruses infect cells lacking cognate receptors by an alternative pathway: implications for retrovirus evolution and xenotransplantation. *J Virol* 78:8868–8877. <http://dx.doi.org/10.1128/JVI.78.16.8868-8877.2004>.
  67. Li J, Akagi K, Hu Y, Trivett AL, Hlynialuk CJ, Swing DA, Volfovsky N, Morgan TC, Golubeva Y, Stephens RM, Smith DE, Symer DE. 2012. Mouse endogenous retroviruses can trigger premature transcriptional termination at a distance. *Genome Res* 22:870–884. <http://dx.doi.org/10.1101/gr.130740.111>.
  68. Jha AR, Pillai SK, York VA, Sharp ER, Storm EC, Wachter DJ, Martin JN, Deeks SG, Rosenberg MG, Nixon DF, Garrison KE. 2009. Cross-sectional dating of novel haplotypes of HERV-K 113 and HERV-K 115 indicate these proviruses originated in Africa before *Homo sapiens*. *Mol Biol Evol* 26:2617–2626. <http://dx.doi.org/10.1093/molbev/msp180>.
  69. Prudhomme S, Oriol G, Malet F. 2004. A retroviral promoter and a cellular enhancer define a bipartite element which controls env ERVWE1 placental expression. *J Virol* 78:12157–12168. <http://dx.doi.org/10.1128/JVI.78.22.12157-12168.2004>.
  70. Vernochet C, Redelsperger F, Harper F, Souquere S, Catzeflis F, Pierron G, Nevo E, Heidmann T, Dupressoir A. 2014. The captured retroviral envelope syncytin-A and syncytin-B genes are conserved in the Spalacidae together with hemotrichorial placentation. *Biol Reprod* 91:148. <http://dx.doi.org/10.1095/biolreprod.114.124818>.
  71. Januszkeski MM, Cannon PM, Chen D, Rozenberg Y, Anderson WF. 1997. Functional analysis of the cytoplasmic tail of Moloney murine leukemia virus envelope protein. *J Virol* 71:3613–3619.
  72. Esnault C, Cornelis G, Heidmann O, Heidmann T. 2013. Differential evolutionary fate of an ancestral primate endogenous retrovirus envelope gene, the EnvV syncytin, captured for a function in placentation. *PLoS Genet* 9:e1003400. <http://dx.doi.org/10.1371/journal.pgen.1003400>.
  73. Imakawa K, Nakagawa S, Miyazawa T. 2015. Baton pass hypothesis: successive incorporation of unconserved endogenous retroviral genes for placentation during mammalian evolution. *Genes Cells* 20:771–788. <http://dx.doi.org/10.1111/gtc.12278>.