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 Mulot, Frédéric Delsuc, Odile Heidmann, Thierry Heidmann, Anne Dupressoir

To cite this version:
Sébastien Malicorne, Cécile Vernochet, Guillaume Cornelis, Baptiste Mulot, 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, American Society for Microbiology, 2016, 90 (18), pp.8132 - 8149. 10.1128/JVI.00483-16. hal-01879328

HAL Id: hal-01879328
https://hal.archives-ouvertes.fr/hal-01879328
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, a Cécile Vernochet, a Guillaume Cornelis, a, b* Baptiste Mulot, c Frédéric Delsuc, d Odile Heidmann, a Thierry Heidmann, a Anne Dupressoir a

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 reinfestation 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 multitransport 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 Dasyopodidae, 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 insertion 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 amion 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-pretreated lungs 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 novemcinct-
<table>
<thead>
<tr>
<th>Species (lineage)</th>
<th>Individual</th>
<th>Origin</th>
<th>Provirus insertion (PCR)</th>
<th>Presence of provirus-related elements (mapping)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Dasypus novemcinctus</em> (U.S.)</td>
<td>T1674</td>
<td>U.S. (MS)</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (U.S.)</td>
<td>T1678</td>
<td>U.S. (TX)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (U.S.)</td>
<td>BM2</td>
<td>U.S. (TX)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (U.S.)</td>
<td>T1829</td>
<td>U.S. (LA)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (U.S.)</td>
<td>ILD128</td>
<td>U.S. (GA)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (U.S.) Female 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (U.S.) Female 2</td>
<td></td>
<td>U.S. (AR)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (U.S.)</td>
<td>BM1</td>
<td>U.S. (TX)</td>
<td>NA</td>
<td>+</td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>JP1</td>
<td>El Salvador (zoo)</td>
<td>NA&lt;sup&gt;a&lt;/sup&gt;</td>
<td>+</td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>JP2</td>
<td>El Salvador (zoo)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>JP3</td>
<td>El Salvador (zoo)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>JP4</td>
<td>El Salvador (zoo)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>JP5</td>
<td>El Salvador (zoo)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>JP6</td>
<td>El Salvador (zoo)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>JP7</td>
<td>El Salvador (zoo)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>JP8</td>
<td>El Salvador (zoo)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>LS1</td>
<td>Mexico (Chiapas)</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>LS2</td>
<td>Mexico (Chiapas)</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>LS3</td>
<td>Mexico (Oaxaca)</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>LS4</td>
<td>Mexico (Chiapas)</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>NP276</td>
<td>Mexico (Oaxaca)</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>YUC1</td>
<td>Mexico (Yucatan)</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>YUC2</td>
<td>Mexico (Yucatan)</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>YUC3</td>
<td>Mexico (Yucatan)</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (East)</td>
<td>T2631</td>
<td>Belize</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (West)</td>
<td>AIR1</td>
<td>Mexico</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (West)</td>
<td>AIR2</td>
<td>Mexico</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus</em> (West)</td>
<td>MCI151</td>
<td>Mexico (Hidalgo)</td>
<td>NA</td>
<td>+</td>
</tr>
<tr>
<td><em>Dasypus pilosus</em></td>
<td>MSB 49990</td>
<td>Peru</td>
<td>NA</td>
<td>+</td>
</tr>
<tr>
<td><em>Dasypus pilosus</em></td>
<td>LSUMZ 21888</td>
<td>Peru</td>
<td>NA</td>
<td>+</td>
</tr>
<tr>
<td>“<em>Dasypus sabanicola</em>”</td>
<td>MSB1</td>
<td>Bolívia (Beni)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td>“<em>Dasypus sabanicola</em>”</td>
<td>MSB2</td>
<td>Bolívia (Beni)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td>“<em>Dasypus sabanicola</em>”</td>
<td>MSB3</td>
<td>Bolívia (La Paz)</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td>“<em>Dasypus sabanicola</em>”</td>
<td>NP751</td>
<td>Peru (Cuzco)</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td>“<em>Dasypus sabanicola</em>”</td>
<td>T1921</td>
<td>Venezuela (Miranda)</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td>“<em>Dasypus sabanicola</em>”</td>
<td>AN1</td>
<td>Bolívia</td>
<td>NA</td>
<td>+</td>
</tr>
<tr>
<td>“<em>Dasypus sabanicola</em>”</td>
<td>AN2</td>
<td>Bolívia</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus sabanicola</em></td>
<td>USNM 372834</td>
<td>Venezuela</td>
<td>NA&lt;sup&gt;a&lt;/sup&gt;</td>
<td>+</td>
</tr>
<tr>
<td><em>Dasypus novemcinctus FG</em></td>
<td>T4069</td>
<td>French Guiana</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus FG</em></td>
<td>T4353</td>
<td>French Guiana</td>
<td>+/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus FG</em></td>
<td>DNO169</td>
<td>French Guiana</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus FG</em></td>
<td>T2863</td>
<td>French Guiana</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus novemcinctus FG</em></td>
<td>T1863</td>
<td>French Guiana</td>
<td>NA</td>
<td>+</td>
</tr>
<tr>
<td><em>Dasypus novemcinctus FG</em></td>
<td>AP207</td>
<td>French Guiana</td>
<td>NA&lt;sup&gt;a&lt;/sup&gt;</td>
<td>+</td>
</tr>
<tr>
<td><em>Dasypus septemcinctus</em></td>
<td>T3002</td>
<td>Argentina</td>
<td>NA</td>
<td>+</td>
</tr>
<tr>
<td><em>Dasypus hybrius</em></td>
<td>T3002</td>
<td>French Guiana</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus hybrius</em></td>
<td>ZVC M2010</td>
<td>Uruguay</td>
<td>NA</td>
<td>+</td>
</tr>
<tr>
<td><em>Dasypus kappleri</em></td>
<td>T4339</td>
<td>French Guiana</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Dasypus kappleri</em></td>
<td>DKA37</td>
<td>French Guiana</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Chaetophractus villosus</em></td>
<td>ChaeVil1</td>
<td>France (zoo)</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Zaedyus pichi</em></td>
<td>T6055</td>
<td>Argentina</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Chlamyphorus truncatus</em></td>
<td>CT1</td>
<td>Argentina</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Cabassous unicinctus</em></td>
<td>T1641</td>
<td>French Guiana</td>
<td>−/−</td>
<td></td>
</tr>
<tr>
<td><em>Tolypeutes matacus</em></td>
<td>Tma TB1</td>
<td>U.S. (zoo)</td>
<td>NA</td>
<td>−</td>
</tr>
<tr>
<td><em>Tolypeutes matacus</em></td>
<td>TolylMat1</td>
<td>France (zoo)</td>
<td>−/−</td>
<td></td>
</tr>
</tbody>
</table>

<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 DnERV(*Env1.1*) sequence.
<sup>e</sup> Illumina reads mapped to the empty locus.
<sup>f</sup> +, Illumina reads mapped to the DnERV(*Env1.1*) sequence; −, no Illumina reads mapped to the DnERV(*Env1.1*) sequence.
tus FG (DNO169) and Cabassous 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 Chaetopterus villosus and Polyphemus maturus 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 Dasypus kappleri, Dasypus hybridus, Záedyus 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 known syncytns, 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-X₅₋₄₋₃-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, 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-Gen1 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 Dasypus 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 30°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 introduced 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”-amplisc) (see Fig. 9B) or the putative 3′-LTR of the provirus within the flanking region for dasy-env1.1 (“env1.1 locus”-amplisc) (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 (1×10⁴ to 2×10⁵ 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 8×10⁵ 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-nls-LacZ (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 (5×10⁴ to 8×10⁵ cells per well) the day before infection, followed by spinoculation at 1,200 X g for 2 h at 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 5×10⁵ 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 ASC2T (see below), a human ASC2T (22), or a human MFS2D (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, synctin-1, or synctin-2 expression vector or an empty vector, each cotransfected with 2.5 μg of the nls-lacZ expression vector. One day after transfection, 3.5×10⁴ 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 ASC2T 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 μg/ml streptomycin, and 100 U/ml penicillin.

**Accession number(s).** The newly determined sequence was deposited in GenBank under accession number KU523789.

**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 Dasypus novemcinctus genome; NCBI, Dasno3v3.0, 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 $\geq 20$ aa located $3'$ to a C-X$_2$-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 (consensus motif of R-X-K/R-R but with a degenerate motif of R-X-K/R-R). 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-env.1 and dasy-env.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-SYSY, 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-SYSY, 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-SYSY cells with the Dasy-Env1.1-expressing vector (supplemented with a β-galactosidase expression vector) led to the formation of multinucleate LacZ-positive (LacZ$^+$) 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 *Dasypus 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_{5,6,7}-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, *Dasypus 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; HTLV2, 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 *Dasypus* 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 Dasypus 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 genes analyzed. (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 β-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+ 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-
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 Dasypus 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 tRNALys1,2 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 substitution 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 proviruses within each subfamily (Fig. 6C). Interestingly, all proviruses share the same be-
taretroviral 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 retrovirus [TvERV(D)]) (46), and mouse (Mus musculus endogenous retrovirus MMERV-B6 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-B6. 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 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- cells were stained in a series of cell lines (see Results), including human HeLa and SH-SYSY 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+ syncytia (red arrows) were detected in human HeLa and SH-SYSY cells but not in A23 cells, with only mononucleated cells being visible by using the empty vector. Original magnification, ×100.](image-url)
ASCT2 (dasyASCT2), the human syncytin-2 receptor MFSD2 (38), or an empty vector (none). One day after transfection, cells were resuspended, and 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), Dasypus 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.

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), Dasypus 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.

Distribution of the DnERV(Env1.1) provirus and related elements in the genus Dasypus. 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 Dasypus (family Dasyopodidae) 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 Dasypus 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 Dasypus 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-
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.
tion from Central America (Mexico, El Salvador, and Belize), 12
samples were heterozygous for the provirus, whereas 6 were neg-
ative, 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 $D.\ novemcinctus$
populations (the latter possibly represents a distinct species [31]),
with 3 and 2 individuals having one copy of the provirus, respec-
tively, and 2 individuals of both species being negative. For the two
most basal species of the genus $Dasyus$ ($D.\ kappleri$ and $D.\ hybri-
dus$), the empty-locus primer pairs amplified the expected ampli-
con in the 3 individuals tested, whereas no 5' or 3' proviral junc-
tion 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

\[FIG 7\] 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 (SDGGGX2DX2R) 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). DnERV, coding sequence reconstituted from the DrERV_216 noncoding Env sequence (GenBank accession number KP175581).

\[FIG 8\] 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 AA866528.1 and AF208161), TVERV(D) (accession numbers AF242745 and AF284693), SMRV (accession numbers NC_001550.1), SRV1 (accession number NT_039167), HERV-K (accession number M11841.1), MPMV (accession number NC_014474.1), MMERV-ß6 (accession number NC_001503.1), enJSRV (accession numbers AAD45226.1 and AAF22166.1), and HIV-1 (accession number K02013).
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 *Dasypus* 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 *Dasypus* (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 *Dasypus* 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_6,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 Dasypus, 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., Euphractus, 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 hoffmannii) (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 Dasypus ancestral lineage between 12 Mya and 45 Mya and (ii) the DnERV-1 family originally inserted into the genome of the least before the divergence between the Dasy-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 reinsertion 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, Das-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 retroviral proteins 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 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 (spermatogonia 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 affinity 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 PMPV (48) as well as for rabbit syncytin-Ory1-ERV, but DrERV, possum Terv(D), and murine MMERV-βδ, 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 Dasypus 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 Dasypus lineage prior to the divergence of D. novemcinctus FG and other Dasypus 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 Dasypus >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 Dasypus 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 humans, 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 Dasypus 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 Dasypus 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 Dasypus 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-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 cytrophoblasts, which normally fuse to form the syncytiotrophoblast (data not shown). Although DnERV(Env1.1) integrated into the Dasypus 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 Dasypus 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’Evolution, 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 Lavialle for comments and critical reading of the manuscript and three anonymous reviewers for helpful suggestions. This work was supported by the CRNS and by grants from the Ligue Nationale contre le Cancer (T.H.; Equipe Labelisée) and the Agence Nationale de la Recherche (ANR) via C-terminal truncation (71). In this respect, ma- and beta-type Env proteins, which can be rendered fusogenic through truncation, was funded by Agence Nationale de la Recherche (ANR) 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).

The funders had no role in study design, data collection and interpretation, and provided no funding for manuscript preparation. 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).

FUNDING INFORMATION

This work, including the efforts of Sebastien Malincorne, Cécile Vernochet, Guillaume Cornélis, Odile Heidmann, Thierry Heidmann, and Anne Dupressoir, was funded by the Agence Nationale de la Recherche (ANR) (RETRO-PLACENTA). The funders had no role in study design, data collection and interpretation, and provided no funding for manuscript preparation.

REFERENCES


Downloaded from http://jvi.asm.org on September 1, 2016 by INIST-CNRS

jvi.asm.org 8147

17. Lavialle C, Cornelis G, Dupressoir A, Essault C, Heidmann O, Vernio-
29. Heidmann, Chiu

31. Gibb, Esnault

32. Taulman, Cornelis

35. Stamatakis

36. Katzourakis

37. Kearse

38. 8148

39. jvi.asm.org


61. Heidmann, Taulman, Cornelis


