

The ~30-million-year-old ERVPb1 envelope gene is evolutionarily conserved among hominoids and Old World monkeys[☆]

Lars Aagaard^{a,1,2}, Palle Villesen^{a,*}, Anders Langfelt Kjeldbjerg^{b,1}, Finn Skou Pedersen^{b,c}

^a Bioinformatics Research Center, University of Aarhus, DK-8000 Aarhus, Denmark

^b Department of Molecular Biology, University of Aarhus, DK-8000 Aarhus, Denmark

^c Department of Medical Microbiology and Immunology, University of Aarhus, DK-8000 Aarhus, Denmark

Received 18 May 2005; accepted 26 August 2005

Available online 2 November 2005

Abstract

Most human endogenous retroviruses (HERVs) are ancient and their genes are rendered nonfunctional by debilitating mutations. One exception is a recently discovered envelope gene located on chromosome 14. This envelope protein was also recently shown to be expressed in various human tissues and to mediate cell–cell fusion *ex vivo*. In this study, we demonstrate that this locus (designated ERVPb1) is preserved in Old World monkeys and that the reading frame is maintained. This is congruent with the entry of the HERV-P(b) group between 27 and 36 million years ago as suggested by long terminal repeat divergence. Although the coding capacity is generally lost in the HERV-IP supergroup, the analysis of nucleotide substitutions, lack of stop codons, and single-nucleotide polymorphisms strongly indicates a selective advantage of the ERVPb1 envelope genes during primate evolution. The purifying selection and tissue-specific expression of the human ERVPb1 envelope gene provide strong evidence of a beneficial role for the host.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Cell fusion; CKS-17; EnvP(b); Envelope protein; HERV; HERV-IP; Primate evolution; Syncytin

Introduction

Almost half of the human genome is repetitive and one of the major repeat classes is human endogenous retroviruses (HERVs). Most comply with the typical proviral organization of an integrated retrovirus, which includes the *gag*, *pol*, and *env* genes flanked by the long terminal repeat (LTR) structures. Altogether, HERVs, solitary LTRs, and related retroelements constitute 8% of the human genome [1]. Most human endogenous retroviruses are ancient and presumably arose as infections of the germline at least 25 million years ago [2]. Copy numbers have increased by re-infection, by intracellular

transposition, or by means of retrotransposon processing *in trans* [3,4] and may range from a few to more than thousands [2].

Most HERV loci are nonfunctional due to the accumulation of debilitating mutations. No replication-competent HERVs have yet been described, although fully intact members of the evolutionarily younger HERV-K group have been reported [5]. Indeed, most of the intact viral open reading frames in the human genome belong to the HERV-K group; however, of the 29 recently annotated full-length envelope genes, half are ancient gammaretroviral loci [6]. An often cited example is ERV3 (a HERV-R member), in which the *gag* and *pol* genes are littered with stop codons, whereas the *env* reading frame is still open [7]. The preservation of an open envelope reading frame in an otherwise degenerate proviral structure has led to speculations about their possible beneficial role for the host. These include (i) preventing retroviral infections by receptor blocking, (ii) providing cells with immunomodulatory properties, or (iii) being used as triggers to provide cell–cell fusion. Syncytin-1, an Env-derived protein from the HERV-W group, provides a striking example of the latter. Syncytin-1 is believed to mediate formation of the syncytiotrophoblast layer during human placenta formation [8,9] and may thus play an essential role in

[☆] Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. DQ247958, DQ247959, DQ247960, DQ247961, and DQ247962.

* Corresponding author. Bioinformatics Research Center, University of Aarhus, Høegh-Guldbergsgade 10, Building 090, DK-8000 Aarhus, Denmark. Fax: +45 89423077.

E-mail address: palle@birc.au.dk (P. Villesen).

¹ Contributed equally to this work.

² Present address: Division of Molecular Biology, Beckman Research Institute, City of Hope, Duarte, CA 91010, USA.

reproduction. A second conserved and fusigenic envelope protein, syncytin-2, from the HERV-FRD group also shows placenta-specific expression and may play a similar cell fusion role [10]. There is also evidence to suggest that syncytin-1 may provide protection against exogenous retroviruses by receptor interference [11].

Expression of HERVs is linked with a number of human diseases including various cancers and a number of autoimmune disorders [12]. Although causal links between disease and HERV activity remain to be established, it is clear from animal models that expression of endogenous retroviral proteins can affect cell proliferation and invoke or modulate immune responses. Recent work by Anthony and colleagues [13] indicates that syncytin-1 may be directly involved in neuronal damage in multiple sclerosis patients by inducing release of cytotoxic factors. HERV envelope proteins have also been ascribed immunoregulatory properties such as the possible superantigenic (SAg) activity of envelopes from HERV-K18 and HERV-W [14,15] and a functional CKS-17-like immunosuppressive motif in HERV-H Env, which prevents tumor clearance in a murine cancer model [16].

Recently, we screened the human genome for retroviral open reading frames and identified a number of intact and almost intact *gag*, *pol*, and *env* genes [6]. Two new groups, dubbed HERV-V and HERV-P(b), emerged from this study, and their expression pattern and fusion properties were recently tested [17]. Interestingly, this study demonstrated that the HERV-P(b) envelope protein mediates cell–cell fusion *ex vivo*. In this work, we demonstrate that the open envelope reading frame at the ERVPb1 locus is fully conserved in Old World monkeys. Comparative analysis of our data strongly suggests that selection has preserved the envelope gene and we

speculate that HERV-P(b) may have played a beneficial role during primate evolution.

Results and discussion

In a recent annotation of open retroviral reading frames in the human genome, we compiled a list of 29 intact or almost intact envelope genes [6]. One of these is a long (664-amino-acid) open reading frame found in a 5.9-kb-long provirus at chromosome 14q32.12 (Fig. 1), which was recently demonstrated to be fusion-competent [17]. Comparisons of the transmembrane (TM) subunit to various envelope sequences show that this provirus is member of the HERV-IP group [17]. This group comprises approximately 35 *pol*-containing members in the human genome based on Southern blot analysis [18]. The envelope gene was dubbed *envP(b)* since the primer-binding site matches tRNA-Pro at 11 of 15 positions and also in order to distinguish it from the distantly related HERV-P group [17]. We propose naming this locus ERVPb1 since the ERV-P(b) group has three additional members in both the human and the chimpanzee genomes (see Supplementary Table 1, Supplementary Fig. 1). Since human ERVPb1 harbors a large 2.8-kb deletion within the *pol* gene, “classical” reverse transcriptase (RT) taxonomy does not apply. However, the highly related human provirus ERVPb2, a presumed predecessor, is seemingly intact, and phylogenetic analysis of conserved motifs within RT confirms classification to the HERV-IP group (data not shown).

The ERVPb1 envelope gene is conserved in Old World monkeys

Using the 2-kb human envelope DNA sequence from ERVPb1 as a query, we searched the available genomes for

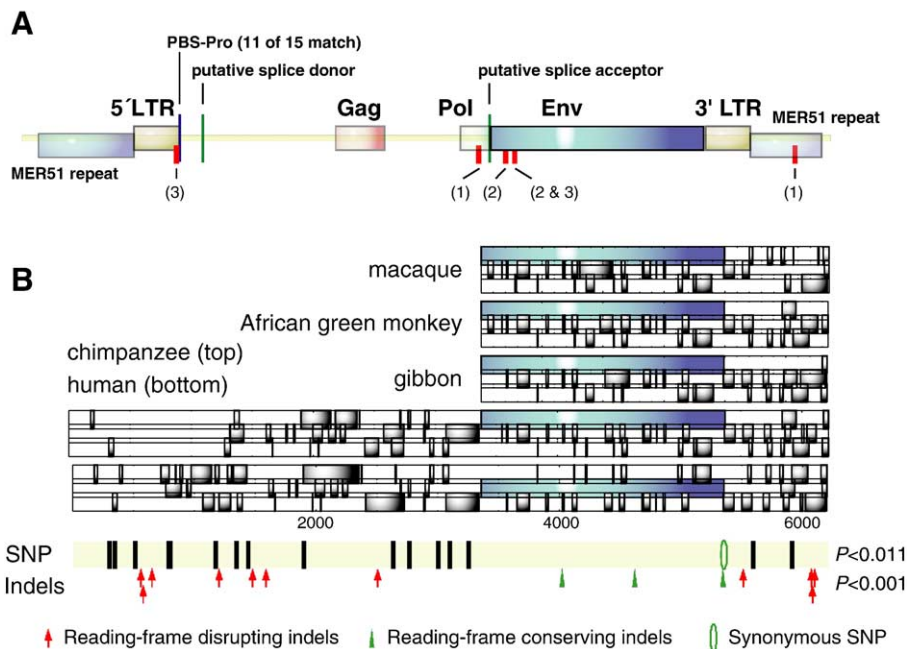


Fig. 1. (A) Proviral organization of HERV-P(b1) provirus. The longest ORF within the *gag*, *pol*, and *env* gene is indicated along with location of primers used for the primate survey (1), expression analysis (2), and detection of a spliced transcript (3). (B) Analysis of start and stop codons in all three reading frames for five species of primates. At the bottom are diagrams of SNPs in the human genome (bars) and the distribution of insertions and deletions (indels) in the alignment (arrows).

homologous genes. No full-length hits above 50–70% identity were detected in the rodent, chicken, or fish genomes. However, in the chimpanzee (*Pan troglodytes*) genome, ERVPb1 is present at the homologous position on chromosome 15. We extended our search by screening a primate DNA panel with PCR primers flanking the ERVPb1 envelope gene (Fig. 1, Table 1). Indeed, we were able to amplify the expected 3-kb amplicon from hominoids (gibbon, GenBank Accession No. DQ247960) and Old World monkeys (African green monkey, GenBank Accession No. DQ247961 and macaque, GenBank Accession No. DQ247962). Interestingly, the envelope reading frame is open in all species examined, and, importantly, several predicted domains as reported by [10], such as the proteolytic cleavage site (RKTR), the CKS-17-like immunosuppressive domain, and the transmembrane region, are fully conserved (Fig. 2, Supplementary Fig. 2). We also note that a CxxCxxxP (CWLC) motif conserved in gammaretroviruses is present in all sequences [19]. Whereas all sequenced ERVPb1 envelope proteins lack a canonical fusion peptide located at the N-terminus of the putative TM subunit, there are two conserved stretches of hydrophobic amino acids located further downstream (Fig. 2). One of these might act as an internal fusion peptide as observed in flaviviruses [20]. Altogether, it is possible that other primate species carry a functional ERVPb1 envelope protein as indicated by the fusiogenic properties of human EnvP(b) [17].

We were unable to amplify the entire ERVPb1 envelope gene from a New World monkey sample (owl monkey, *Aotus trivirgatus*); however, locus-specific intra-envelope primers [as also used for quantitative reverse transcription-polymerase chain reaction (RT-PCR) analysis; Fig. 1] did amplify the expected 100-bp envelope fragment from owl monkey DNA. Comparative sequence analysis confirms that this fragment most likely was amplified from ERVPb1 and we saw no stop codons or frameshift mutations (data not shown). However, limited sequence variation among the human and chimpanzee ERV-P(b) members in this region and lack of flanking DNA sequences make definitive locus assignment difficult. We also attempted to PCR-amplify the long 3-kb and the short 100bp envelope amplicon from various prosimian lemur species (ring-tailed lemur, *Lemur catta*, and gray mouse lemur, *Microcebus murinus*). All of these samples consistently showed no amplicon, indicating that the ERV-P(b1) integration happened

after the catarrhine–strepsirrhine split. Although caution should be taken in interpreting these negative results, the presence of ERV-P(b1) in catarrhines is in accordance with previously reported results for other HERV-IP members using probe hybridization techniques [18,21].

The ERVPb1 locus is 27–36 MY old

Using a divergence time of 20–25 MYA for the hominoid–cercopithecoid split [22], we estimated the rate of evolution for the 3′ LTR region to 1.3×10^{-9} to 1.7×10^{-9} substitutions \times position⁻¹ \times million year⁻¹ (Supplementary Table 2). Assuming a similar, clock-like rate of evolution between the 5′ and the 3′ LTR regions, we can estimate that the ERVPb1 provirus entered the genome approximately 27–36 MYA (Fig. 3, Table 2). This is in accordance with the lack of ERV-P(b1) in prosimians as suggested by our PCR survey and also the absence of ERV-IPs in prosimians as shown by a recent microarray-based analysis [21]. The remaining three HERV-P(b) members show a similar level of divergence between the 5′ and 3′ LTRs (ranging from 10 to 12.5%, corrected), indicating that this subgroup entered the genome within a relatively short period of time. This is further substantiated by the narrow branching order of the four members of the HERV-P(b) clade within the HERV-IP group (Supplementary Fig. 1).

Evidence for selection within the ERVPb1 env gene

Within the coding regions only the *env* gene is intact. The majority of the *pol* gene has been lost by a 2.8-kb deletion and there is a density of >1 stop codon per 20 codons within the *gag* gene and only residual reading frames remain (Fig. 1). The four HERV-P(b) members contain only a single intact gene (the ERVPb1 *env*), which surprisingly was found to be intact in all species tested.

To further validate this strong indication of purifying selection, we applied normal K_a/K_s tests for the entire *env* gene. These did not indicate a clear pattern of significant deviation from a K_a/K_s ratio \sim 1, but subsequent site-specific analysis revealed varying K_a/K_s ratios over the gene (Fig. 2). For comparison, we also applied the classical K_a/K_s ratio test to the syncytin-1 locus [23], which similarly shows no overall pattern of significant purifying selection, whereas the syncytin-2 locus [10] shows strong significance of selection (data not shown).

The strongest evidence of selection within the ERVPb1 *env* gene comes from the patterns of human single-nucleotide polymorphism (SNP) variation and insertions and deletions between primate species. By mapping SNP polymorphisms onto the human sequence, we observed that only 1 of 17 SNPs falls within the envelope open reading frame (ORF) and that single SNP is synonymous (Fig. 1). Assuming random distribution of SNPs, the binomial probability of observing this pattern (16 outside, 1 inside; ignoring nonsynonymous/synonymous information) is significant ($P = 0.011$).

Furthermore, the alignment contains 14 insertion/deletions events. Three of these do not disrupt the reading frame (i.e., indel length = $n \times 3$) and are located in the envelope region.

Table 1
Sequence similarity of the ERVPb1 envelope gene in five species of primates

	Human	Chimpanzee	Gibbon	African green monkey	Macaque
Human		0.984	0.893	0.855	0.854
Chimpanzee	0.989		0.887	0.854	0.852
Gibbon	0.951	0.948		0.848	0.848
African green monkey	0.926	0.925	0.918		0.960
Macaque	0.923	0.923	0.917	0.981	

Note. Protein identities are depicted in the upper triangle and DNA identities in the lower triangle.

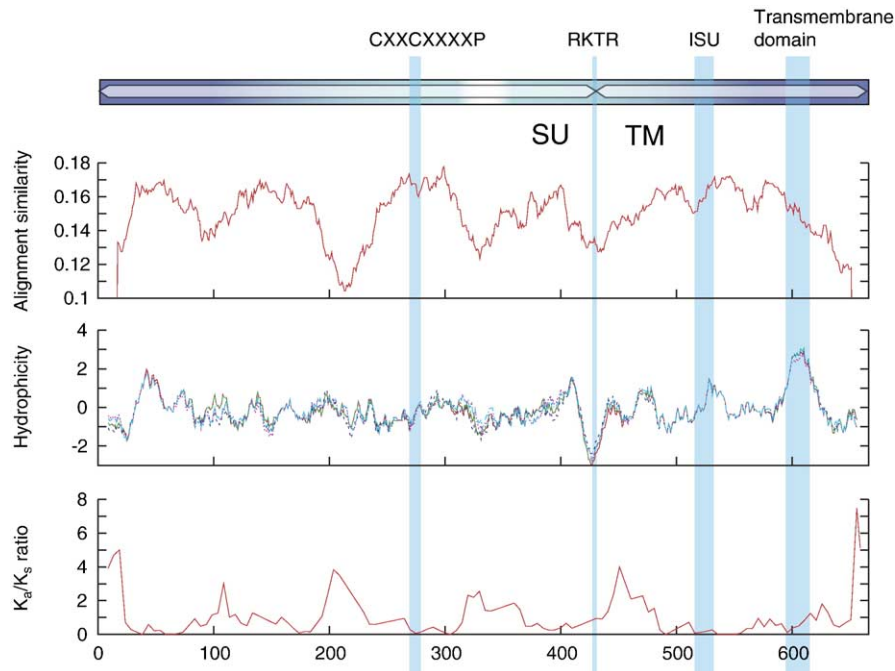


Fig. 2. Annotation and sliding-window analysis of alignment similarity, hydrophobicity, and K_a/K_s ratios for the aligned ERVPb1 envelope genes.

The remaining 11 all do disrupt the reading frame and all of them are found outside the envelope region (Fig. 1). Assuming random distribution of indels (i.e., assuming neutral selection), the binomial probability of observing the 11 disrupting indels outside and 0 disrupting indels inside the envelope ORF is again significant ($P = 0.00081$).

Expression levels in human tissue samples

The recent report by [17] demonstrates that *envP(b)* is widely expressed in human tissues. Our data (Supplementary Fig. 3) confirm this ubiquitous expression, although there may be minor differences in the relative expression levels between tissues compared to the data from Blaise et al. [17]. Generally,

envP(b) expression is low but consistent in a wide range of tissues and cell lines (Supplementary Fig. 3). When we compare the expression to syncytin-1 levels in placental RNA, *envP(b)* expression is roughly 100-fold lower (data not shown). Interestingly, a closely related *env*-deficient element (HERV-IP-T47D, Supplementary Table 1) is expressed and packaged into virus particles that are spontaneously released from mammary carcinoma-derived cells [18]. Since HERV-P(b1) is the only HERV-IP member with an intact *env* gene, it is possible that EnvP(b) could be involved in such tumor-associated particle formation. As depicted in Fig. 1, the human ERVPb1 provirus is flanked by two 425-bp LTRs. When scanned for possible enhancer and promoter motifs, it seems likely that the ERVPb1 5' LTR might still direct transcriptional

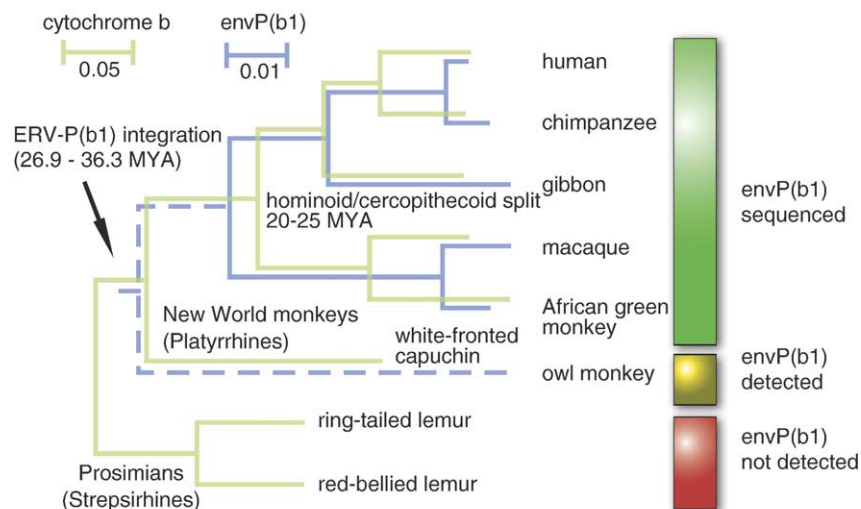


Fig. 3. Maximum-likelihood tree of ERV-P(b1) envelope DNA sequences superimposed on primate phylogeny using cytochrome b sequences. The time estimate for the hominoid–cercopithecoid split was used to calculate the age of the ERVPb1 locus (see Table 2 and Supplementary Table 2).

Table 2

Age of ERVPb1 integration using intraspecific 5'–3' LTR divergences from human and chimpanzee genome sequences

	LTR divergence	Substitution rate	Age estimate
Human	0.0990	1.36×10^{-3} to 1.70×10^{-3}	29.07–36.34
Chimpanzee	0.0882	1.31×10^{-3} to 1.64×10^{-3}	26.95–33.69

Note. Substitution rates are based on the interspecific LTR divergence (Supplementary Table 2) and a divergence time of the hominoid–cercopithecoid split of 20–25 MYA [22].

activity. In particular, we find an intact TATA-box (Supplementary Fig. 4). To address the question of whether the envelope-specific transcript, detected by RT-qPCR primers by us and by Blaise et al. [17], might stem from a spliced and hence in vivo translation-competent RNA, we performed RT-qPCR using a primer set spanning the viral leader and the envelope gene. Indeed, we detect a 500-bp band, which is consistent with splicing from a putative splice-donor site in the leader region just upstream of the first AUG site in *env*.

Concluding remarks

We have amplified and sequenced the *envP(b)* gene located on chromosome 14q32.12 (previously named ZFERV-like Env) from hominoids (gibbon) and Old World monkeys (African green monkey and macaque). We propose naming this locus ERVPb1. Our data surprisingly demonstrate that the ERVPb1 envelope reading frame is open and intact in all species examined (Fig. 1), and a detailed comparative analysis suggests that parts of this gene have been subjected to purifying selection to maintain the amino acid composition ($K_a/K_s < 1$, Fig. 2). Some of the conserved regions coincide with known motifs from infectious retroviral envelope proteins such as the CxxCxxxP region, the immunosuppressive motif, and the transmembrane region. Taken together with the recently described ability of human EnvP(b) to fuse cells *ex vivo* [17], it is plausible that the primate EnvP(b) proteins may also mediate cell–cell fusion or even infectivity. We are currently testing this hypothesis. Moreover, the deviation from neutral selection at this locus is clearly demonstrated by significant lack of insertions and deletions (indels) and SNPs within the envelope gene compared to the flanking proviral sequence (Fig. 1).

Using a locus-calibrated molecular clock and by assuming identical LTR at the time of integration, the level of divergence between the flanking LTRs leads to an age estimate of ERVPb1 of approximately 30 MY. This suggests that ERVPb1 may also be present in New World monkeys. Indeed, we do detect a 100-bp ERV-P(b) fragment in owl monkeys. This observation warrants further investigations into the presence or absence of an intact envelope gene in New World monkeys.

We confirm the data from Blaise et al. [17] that the human ERVPb1 gene is expressed in a tissue-specific manner. The positive expression in placenta is particularly interesting since two other fusogenic envelope proteins, syncytin-1 and -2, are also expressed in this tissue. Syncytin-1 and -2 have been suggested to play a role in placenta formation by mediating

fusion of trophoblast cells [8–10]. Thus, considering the strong evolutionary conservation of ERVPb1, it is possible that EnvP(b) plays a similar physiological role. However, as noted by [17], the ubiquitous pattern of expression does not resemble the placenta-specific expression observed for the two syncytin genes. Furthermore, the levels of expression in the placenta are approximately 100-fold lower than that observed for syncytin-1. Our data suggest that the human ERVPb1 gene is in fact most highly expressed in the thymus. The physiological implications of thymic expression, if any, are not clear, although an immunological component could be suggested.

Almost all of the intact non-HERV-K retroviral reading frames present in the human genome are envelope proteins [6]. A physiological role as a trigger for cell–cell fusion has been suggested for the intact HERV-W and HERV-FRD envelope genes (syncytin-1 and -2), as discussed above. Expression of an endogenous envelope protein may also protect against a related disease-causing virus by receptor blockage. Wild mice expressing the Fv4 envelope gene are thus less susceptible to infection by murine leukemia virus [24]. Syncytin-1 expression has in fact been shown to provide infection resistance toward the avian spleen necrosis virus [11], and a similar role could apply to other envelope loci that presumably still recognize a cellular receptor such as Syncytin-2 and EnvP(b). Alternative functional roles for endogenous envelope proteins that may turn out to be unrelated to retroviral biology may also await discovery.

The data presented here show that the intact EnvP(b) ORF is not only conserved in humans, but has remained intact for millions of years in all species studied. The simplest and most likely explanation is that it is a functional gene. Altogether, our evolutionary analysis provides compelling evidence that an intact EnvP(b) gene plays a beneficial role in simians, but exactly how this effect is mediated must await further investigations.

Materials and methods

DNA samples, PCR amplification, and sequencing

DNA from gibbon (*Hylobates gabriellae*) and macaque (*Macaca nemestrina*) was a gift from Max Planck Institute for Evolutionary Anthropology (Leipzig), tissue from ring-tailed lemur (*Lemur catta*) was a gift from Living United/Ebeltoft Zoo & Safari (DNA was isolated using a QIAGEN tissue isolation kit), DNA from gray mouse lemur (*Microcebus murinus*) was kindly donated by Tina Fredsted, DNA from owl monkey (*Aotus trivirgatus*) was purchased from ECACC (OMK 637-69), and DNA from African green monkey (*Cercopithecus aethiops*) was isolated from COS cells (using Invitrogen DNazol). All primate DNA samples were verified by amplification and direct sequence analysis of the cytochrome oxidase I (COI) gene. Sequences from human (*Homo sapiens*) and chimpanzee (*Pan troglodytes*) were extracted from the ENSEMBL genome database (<http://www.ensembl.org>). Primers were designed to amplify EnvP(b1) and a part of the 3' flanking region including the 3' LTR region (No. 1, Table 3). The PCR was performed by the use of rTth DNA Polymerase, XL (Applied Biosystem) in 100- μ l reaction using 300 nM primer, 0.2 mM dNTP, 5 μ l of Mg(OAc)₂, 150 ng of DNA, and buffers by the supplier. The PCR program was as follows: 1 cycle (94°C for 1 min), 35 cycles (94°C for 1 min; 60°C for 1 min; 72°C for 5 min), and 1 cycle (72°C for 10 min). Resulting PCR products (3 kb) were directly sequenced from a 5- μ l PCR mixture using the ET sequencing kit (Amersham) according to the manufacturer (details for sequence primers are available from the authors upon request) and analyzed on an ABI3100 (Applied Biosystems). All ERVPb1 sequences have

Table 3
Primers used for the ERVPb1 primate survey and human expression analysis

Name	Sequence (5'–3')
No. 1 forward	GGTCCCTGATCCCTACATGTGGCACAGCTCAGAGC
No. 1 reverse	CCAAAGAGTGTCTGGAGCCAAACAGTACCAAAGGAGG
No. 2 forward	TCTTTCCTTGTGCTCTTGCTAGCCA
No. 2 reverse	GGCTTTGGAGATCCCCACAAGGTA
No. 3 forward	TTTCACTTCACAGCACTGACAATTTGT
No. 3 reverse	Identical to No. 2 reverse
COI forward	CCCCGAATAAACAATATAAGCTTCTG
COI reverse	AAAAATCAGAATAGGTGTTGGTATAG
β -Actin forward	CCGCCAGCTCACCATGGATGAT
β -Actin reverse	CTTCTGACCCATGCCACCA
GAPDH forward	GTCAGCCGCATCTTCTTTTG
GAPDH reverse	GCGCCCAATACGACCAAATC
Syncytin-1 forward	CCCCATCGTATAGGAGTCTT
Syncytin-1 reverse	CCCCATCAGACATACCAGTT

been deposited with the GenBank Data Library under the following accession numbers: (human ERVPb1, GenBank Accession No. DQ247958), (chimpanzee ERVPb1, GenBank Accession No. DQ247959), (gibbon ERVPb1, GenBank Accession No. DQ247960), (African green monkey ERVPb1, GenBank Accession No. DQ247961), and (macaque ERVPb1, GenBank Accession No. DQ247962).

Expression analysis using quantitative RT-PCR

RNA samples from various human cell lines were isolated using TriZol Reagent (Invitrogen). RNA from 24 different human tissues (Total RNA Master Panels II) was obtained from BD Biosciences. All RNA samples were treated with TURBO DNase 1U (Ambion) for 30 min at 37°C, phenol/chloroform extracted, ethanol precipitated, and dissolved in nuclease-free water (Ambion). One microgram of RNA was reverse-transcribed using the M-MLV reverse transcriptase (Invitrogen) using random hexamer primers. The resulting cDNA was diluted so that 1 μ l corresponded to 17 ng of total RNA. RT-qPCR analysis was performed on a MX4000 (Stratagene) using 10 μ l cDNA, 100 nM primers (No. 2, Table 3) and SYBR Green (Stratagene) in a 25- μ l reaction according to the manufacturer's instructions. The PCR program was as follows: 1 cycle (94°C for 10 min) and 40 cycles (94°C for 0.5 min; 60°C for 1 min; 72°C for 0.5 min). As an internal control, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and β -actin genes were also amplified (data not shown). The relative expression in each sample was calculated with respect to a standard calibration curve (a dilution series of genomic DNA). Each sample was analyzed twice, and the PCR products were sequenced to verify the target specificity. To detect a spliced envelope transcript, we amplified an ~500-bp product using a forward primer upstream of the putative splice-donor site and a reverse primer within the envelope reading frame (No. 3) from the placental cDNA sample.

Evolutionary analysis

Sequences were aligned using ClustalW [25] and alignment similarity plots and protein hydrophobicity plots were created using the EMBOSS package [26]. The K_a/K_s ratio data were calculated using PAML [27] and a sliding window of 17 aa. To test for evolutionary conservation of the ORF, we downloaded human SNP data from ENSEMBL (<http://www.ensembl.org>) and identified reading-frame-disrupting indels in the alignment. Assuming a random distribution of SNPs and insertion/deletions, we calculated the P values as the binomial probability of the observed SNP and insertion/deletion patterns. Promoter and enhancer predictions for the LTR regions were carried

out using TFSEARCH (<http://www.cbrc.jp/research/db/TFSEARCH.html>) and ALIBABA2 (<http://www.alibaba2.com>).

ML phylogenies were created using PHYLIP [28] with default parameters. The ERVPb1 envelope tree was based on the envelope DNA (alignment length: 2000 bp). A primate tree of the cytochrome b DNA was also constructed for comparison with the primate phylogeny. The HERV-IP envelope tree was based on a manually inspected envelope DNA (alignment length: 727 bp) to accommodate alignment to the outgroup HERV-ADP. References to all sequences are provided in Supplementary Table 1.

Substitution rates were estimated using data from five species of primates and using a divergence time of 20–25 MYA of the hominoid–cercopitheoid split [22]. The HERV integration time was then calculated from the LTR divergence (since the LTRs were assumed to be identical when the HERV was integrated) and the estimated substitution rate (Kimura 2 parameter).

Acknowledgments

This work was supported by the Danish Cancer Society, the Danish Technical Research Council, the Karen Elise Jensen Foundation, and Aarhus University Research Foundation. The authors thank Birgit Nickle (Max Planck Institute for Evolutionary Anthropology), Jesper Stagegaard (Living United/Ebeltoft Zoo and Safari), and Tina Fredsted (University of Aarhus) for providing primate DNA samples. The authors also thank two anonymous reviewers for help in improving the manuscript. This study was conceived by all authors. L.A. coordinated the study, drafted the manuscript, and participated in the statistical and evolutionary analysis. P.V. performed sequence manipulations and statistical and evolutionary analyses, prepared the figures, and participated in drafting the manuscript. A.L.K. performed all experimental work described in this study and participated in drafting the manuscript. F.S.P. coordinated the study and participated in critical revision of the manuscript. All authors read and approved the final manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.ygeno.2005.08.011](https://doi.org/10.1016/j.ygeno.2005.08.011).

References

- [1] IHGSC, Initial sequencing and analysis of the human genome, *Nature* 409 (2001) 860–921.
- [2] J. Boeke, J. Stoye, Retrotransposons, endogenous retroviruses, and the evolution of retroelements, In *Retroviruses* (1997) 343–435.
- [3] R. Belshaw, A. Katzourakis, J. Paces, A. Burt, M. Tristem, High copy number in human endogenous retrovirus (HERV) families is associated with copying mechanisms in addition to re-infection, *Mol. Biol. Evol.* 22 (4) (2005) 814–817.
- [4] A. Pavlicek, J. Paces, D. Elleder, J. Hejnar, Processed pseudogenes of human endogenous retroviruses generated by LINEs: Their integration, stability, and distribution, *Genome Res.* 12 (3) (2002) 391–399.
- [5] G. Turner, M. Barbulescu, M. Su, M. Jensen-Seaman, K. Kidd, J. Lenz, Insertional polymorphisms of full-length endogenous retroviruses in humans, *Curr. Biol.* 11 (2001) 1531–1535.
- [6] P. Villesen, L. Aagaard, C. Wiuf, F. Pedersen, Identification of endogenous retroviral reading frames in the human genome, *Retrovirology* 1 (1) (2004) 32.
- [7] M. Cohen, M. Powers, C. Oconnell, N. Kato, The nucleotide-sequence

- of the Env gene from the human provirus ERV3 and isolation and characterization of an ERV3-specific cDNA, *Virology* 147 (1985) 449–458.
- [8] J.L. Blond, D. Lavillette, V. Cheynet, O. Bouton, G. Oriol, S. Chapel-Fernandes, B. Mandrand, F. Mallet, F.L. Cosset, 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 (7) (2000) 3321–3329.
- [9] S. Mi, X. Lee, X. Li, G.M. Veldman, H. Finnerty, L. Racie, E. LaVallie, X.Y. Tang, P. Edouard, S. Howes, et al., Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis, *Nature* 403 (6771) (2000) 785–789.
- [10] S. Blaise, N. de Parseval, L. Benit, T. Heidmann, Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution, *Proc. Natl. Acad. Sci. USA* 100 (22) (2003) 13013–13018.
- [11] V.G. Ponferrada, B.S. Mauck, D.P. Wooley, The envelope glycoprotein of human endogenous retrovirus HERV-W induces cellular resistance to spleen necrosis virus, *Arch. Virol.* 148 (4) (2003) 659–675.
- [12] R. Lower, The pathogenic potential of endogenous retroviruses: Facts and fantasies, *Trends Microbiol.* 7 (1999) 350–356.
- [13] J.M. Antony, G. van Marle, W. Opii, D.A. Butterfield, F. Mallet, V.W. Yong, J.L. Wallace, R.M. Deacon, K. Warren, C. Power, Human endogenous retrovirus glycoprotein-mediated induction of redox reactants causes oligodendrocyte death and demyelination, *Nat. Neurosci.* 7 (10) (2004) 1088–1095.
- [14] B. Conrad, R. Weissmahr, J. Boni, R. Arcari, J. Schupbach, B. Mach, A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes, *Cell* 90 (1997) 303–313.
- [15] H. Perron, E. Jouvin-Marche, M. Michel, A. Ounanian-Paraz, S. Camelo, A. Dumon, C. Jolivet-Reynaud, F. Marcel, Y. Souillet, E. Borel, et al., Multiple sclerosis retrovirus particles and recombinant envelope trigger an abnormal immune response in vitro, by inducing polyclonal Vbeta16 T-lymphocyte activation, *Virology* 287 (2001) 321–332.
- [16] M. Mangeney, N. de Parseval, G. Thomas, T. Heidmann, The full-length envelope of an HERV-H human endogenous retrovirus has immunosuppressive properties, *J. Gen. Virol.* 82 (Pt. 10) (2001) 2515–2518.
- [17] S. Blaise, N. de Parseval, T. Heidmann, Functional characterization of two newly identified Human Endogenous Retrovirus coding envelope genes, *Retrovirology* 2 (1) (2005) 19.
- [18] W. Seifarth, C. Baust, U. Schon, A. Reichert, R. Hehlmann, C. Leib-Mosch, HERV-IP-T47D, a novel type C-related human endogenous retroviral sequence derived from T47D particles, *AIDS Res. Hum. Retroviruses* 16 (5) (2000) 471–480.
- [19] A. Pinter, R. Kopelman, Z. Li, S.C. Kayman, D.A. Sanders, Localization of the labile disulfide bond between SU and TM of the murine leukemia virus envelope protein complex to a highly conserved CWLC motif in SU that resembles the active-site sequence of thiol–disulfide exchange enzymes, *J. Virol.* 71 (10) (1997) 8073–8077.
- [20] S.L. Allison, J. Schalich, K. Stiasny, C.W. Mandl, F.X. Heinz, Mutational evidence for an internal fusion peptide in flavivirus envelope protein E, *J. Virol.* 75 (9) (2001) 4268–4275.
- [21] A.D. Greenwood, A. Stengel, V. Erfle, W. Seifarth, C. Leib-Mosch, The distribution of pol containing human endogenous retroviruses in non-human primates, *Virology* 334 (2) (2005) 203–213.
- [22] A.D. Yoder, Z. Yang, Estimation of primate speciation dates using local molecular clocks, *Mol. Biol. Evol.* 17 (7) (2000) 1081–1090.
- [23] B. Bonnaud, O. Bouton, G. Oriol, V. Cheynet, L. Duret, F. Mallet, Evidence of selection on the domesticated ERVWE1 env retroviral element involved in placentation, *Mol. Biol. Evol.* 21 (10) (2004) 1895–1901.
- [24] M. Gardner, C. Kozak, S. O'Brien, The Lake Casitas wild mouse: Evolving genetic resistance to retroviral disease, *Trends Genet.* 7 (1991) 22–27.
- [25] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22 (22) (1994) 4673–4680.
- [26] P. Rice, I. Longden, A. Bleasby, EMBOSS: The European Molecular Biology Open Software Suite, *Trends Genet.* 16 (6) (2000) 276–277.
- [27] Z. Yang, PAML: A program package for phylogenetic analysis by maximum likelihood, *Comput. Appl. Biosci.* 13 (5) (1997) 555–556.
- [28] J. Felsenstein, PHYLIP (Phylogeny Inference Package). Version 3.6. Department of Genome Sciences, University of Washington, Seattle, 2004.