

Identification of Novel Endogenous Betaretroviruses Which Are Transcribed in the Bovine Placenta[∇]

Kenji Baba,^{1†} Yuki Nakaya,^{1†} Takayuki Shojima,¹ Yoshikage Muroi,² Keiichiro Kizaki,³ Kazuyoshi Hashizume,³ Kazuhiko Imakawa,² and Takayuki Miyazawa^{1*}

Laboratory of Signal Transduction, Department of Cell Biology, Institute for Virus Research, Kyoto University, 53 Shogoin-Kawaracho, Sakyo-ku, Kyoto-city, Kyoto 606-8507,¹ Laboratory of Animal Breeding, Graduate School of Agricultural and Life Sciences, the University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657,² and Laboratory of Veterinary Physiology, Department of Veterinary Medicine, Faculty of Agriculture, Iwate University, 3-18-8 Ueda, Morioka, Iwate 020-8550,³ Japan

Received 8 June 2010/Accepted 3 November 2010

Sequences of retroviral origin occupy approximately 10% of mammalian genomes. Various infectious endogenous retroviruses (ERVs) and functional retroviral elements have been reported for several mammals but not cattle. Here, we identified two proviruses, designated bovine endogenous retrovirus K1 (BERV-K1) and BERV-K2, containing full-length envelope (*env*) genes in the bovine genome. Phylogenetic analysis revealed that they belong to the genus *Betaretrovirus*. By reverse transcription (RT)-PCR, both BERV-K1 and -K2 *env* mRNAs were detected in the placenta and cultured bovine trophoblast cells. Real-time RT-PCR analysis using RNAs isolated from various bovine tissues revealed that BERV-K1 *env* mRNA was preferentially expressed in the placenta. Moreover, we also found the expression of doubly spliced transcripts, named the *REBK1* and *REBK2* genes. Both the *REBK1* and *REBK2* proteins have motifs for a putative nuclear localization signal and a nuclear export signal. *REBK1* and *REBK2* fused with green fluorescent proteins were localized mainly in the nuclei when they were expressed in bovine and porcine cells. In the *env* and 3' long terminal repeats of BERV-K1 and -K2, we found regulatory elements responsible for the splicing and transport of viral RNAs and/or translation of the *env* genes. Although we have not identified the expressed Env proteins in bovine tissues, these data suggest that both BERV-K1 and BERV-K2 express Env proteins and that these proteins may have physiological functions *in vivo*.

Endogenous retroviruses (ERVs) are retroviruses which have integrated their proviral genome into host germ line cells, and evidently, they have been found to occupy approximately 10% of the host genome in mammals (14, 33). Although most ERVs have been inactivated by mutations and/or deletions, a few open reading frames (ORFs) of ERVs are still active, leading to viral protein expression in their hosts. Interestingly, the envelope (Env) proteins of ERVs have been reported to be preferably expressed in the placenta and are involved in placental morphogenesis in various mammals, including humans, sheep, mice, and rabbits (2, 7, 8, 9, 11, 23). The Env protein of human endogenous retrovirus W (HERV-W), also named syncytin-1, is expressed in trophoblast cells and induces cell fusion *in vitro* (23). The Env protein of endogenous Jaagsiekte sheep retrovirus (JSRV) (enJSRV) is expressed in trophoblast cells and is essential for the growth and differentiation of the cells but not cell fusion in the peri-implantation conceptus (7). Homozygous syncytin-A-null mouse embryos were shown to die *in utero* between 11.5 and 13.5 days of gestation, and the syncytin-A-deficient placenta results in the specific disruption of the architecture of the syncytiotrophoblast-containing laby-

rinth, with trophoblast cells failing to fuse into an interhemal syncytial layer (9). This evidence suggests the coevolution of ERVs and mammals following the invasion of ancestral retroviruses in mammals.

Cattle (family *Bovidae*, genus *Bos*) are one of the most common farm animals belonging to order Artiodactyla, as are pigs (family *Suidae*, genus *Sus*) and sheep (family *Bovidae*, genus *Ovis*). Porcine endogenous retroviruses (PERVs) and enJSRV, which are unique ERVs in the genera *Sus* and *Ovis*, respectively, have been identified and well characterized (1, 6, 17, 29, 30, 36). On the other hand, only a few ERV elements in cattle have been reported (10, 25, 34, 35). To our knowledge, bovine ERV (BERV) elements, which contain intact retroviral ORFs, have not been reported.

In this study, we identified two novel ERV proviruses, designated BERV-K1 and -K2, containing full-length *env* ORFs in the bovine genome by *in silico* analyses. The BERV-K2 provirus on chromosome 24 contained *gag*, *pro*, and *pol* ORFs in addition to the *env* ORF. Both BERV-K1 and -K2 *env* mRNAs were expressed in the placenta *in vivo* and in a bovine trophoblast cell line. Furthermore, we identified alternative ORFs which were processed by the double splicing of the *env* mRNA. We also examined the regulatory mechanisms of BERV-K1 and -K2 *env* mRNAs.

MATERIALS AND METHODS

Database screening and *in silico* sequence analyses. Retroviral *env* sequences homologous to transmembrane (TM) Env sequences of known ERVs, including PERV (GenBank accession number CAA72927), enJSRV (accession number

* Corresponding author. Mailing address: Laboratory of Signal Transduction, Department of Cell Biology, Institute for Virus Research, Kyoto University, 53 Shogoin-Kawaracho, Sakyo-ku, Kyoto-city, Kyoto 606-8507, Japan. Phone and fax: 81-75-751-4814. E-mail: takavet@goo.jp.

† K.B. and Y.N. equally contributed to this work.

∇ Published ahead of print on 17 November 2010.

AAF22166), HERV-W (accession number AAF28334), syncytin-A (accession number AAW62446), and syncytin-B (accession number NP775596), were searched by using the TBLASTN program for bovine genome resources at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/projects/genome/guide/cow/>). BLAST searches were then repeated using positive sequences. Finally, a 2.5-kb region encompassing 2 kb upstream and 500 bp downstream of the test sequence was extracted from positive sequences, and only those containing an ORF of >1.5 kb were selected and analyzed for characteristic features of retroviral Env. When an intact *env* ORF was found, upstream and downstream sequences were further analyzed for other ORFs and 5' and 3' long terminal repeats (LTRs). Primer binding sites (PBSs) were predicted by using the Genomic tRNA Database (<http://lowelab.ucsc.edu/GtRNAdb/>). Protein domains and motifs were searched by using Prosite (<http://www.expasy.org/prosite/>). Cleavage sites of the signal peptide of Env were predicted by using PrediSi (<http://www.predisi.de/index.html>). Hydrophobicity profiles were determined by using PEPWINDOW (<http://www.cbib.u-bordeaux2.fr/pise/pepwindow.html>). A phylogenetic tree was constructed from the alignment using the neighbor-joining program within CLUSTALW (<http://align.genome.jp/>). The nuclear localization signal (NLS) and nuclear export signal (NES) were predicted by using WoLF PSORT (<http://wolfpsort.org/>) and the NetNES 1.1 server (<http://www.cbs.dtu.dk/services/NetNES/>), respectively.

Animals and tissue collection. Bovine tissues (placenta, lung, liver, spleen, heart, muscle, kidney, skin, testis, and endometrium [estrus cycle]) were collected from 3 or 6 Holstein cows. The placenta was collected on day 150 of gestation and divided into cotyledonary villi (COT), intercotyledonary villi (ICOT), caruncular (CAR), and intercaruncular (ICAR). Placenta ($n = 3$), endometrium ($n = 6$), and other tissues ($n = 3$) were collected independently. The collected samples were stored at -80°C prior to RNA extraction. All procedures for these animal experiments were carried out in accordance with guidelines approved by the Animal Committee of Iwate University and the National Institute of Agrobiological Sciences for the use of experimental animals.

Cell lines. Human embryonic kidney 293T (HEK293T) cells, MDBK cells (bovine kidney cells), and PK-15 cells (porcine kidney cells) were grown in Dulbecco's modified Eagle's medium (DMEM) (Sigma, St. Louis, MO) supplemented with 10% heat-inactivated fetal calf serum (FCS), 100 IU/ml penicillin, and 100 $\mu\text{g}/\text{ml}$ streptomycin at 37°C in a humidified atmosphere of 5% CO_2 in air. The establishment of BT-1 cells (bovine trophoblast cells) was described previously (32). The cells were maintained in DMEM-nutrient mixture F-12 Ham (Sigma) supplemented with 10% FCS and antibiotics. Plat-E cells (an ecotropic murine leukemia virus [MLV]-based packaging cell line) (24) were cultured in DMEM supplemented with 10% heat-inactivated FCS, 100 IU/ml penicillin, 100 $\mu\text{g}/\text{ml}$ streptomycin, 1 $\mu\text{g}/\text{ml}$ puromycin, and 10 $\mu\text{g}/\text{ml}$ blasticidin at 37°C in a humidified atmosphere of 5% CO_2 in air.

Construction of expression plasmids. Expression plasmids encoding REBK1 (pREBK1GFP) and REBK2 (pREBK2GFP) fused with a green fluorescent protein (GFP) at the C terminus were constructed by using pAcGFP1-N1 (Clontech Laboratories, Palo Alto, CA) as a backbone. *REBK1* and *REBK2* cDNAs were each ligated into pAcGFP1-N1, which was digested with EcoRI and SmaI. *GFP*, *REBK1-GFP*, and *REBK2-GFP* gene sequences were digested with EcoRI and NotI from pAcGFP1-N1, pREBK1GFP, and pREBK2GFP, respectively, and the fragments were then ligated into the pLPCX retroviral vector (Clontech) to generate pseudotype viruses containing GFP, REBK1GFP, or REBK2GFP, respectively, in Plat-E cells. N-terminal hemagglutinin (HA) epitope-tagged expression plasmids pHAREBK1, pHAREBK2, pHAEnvK1, pHAEnvK2, pHAEnvK13'LTR, and pHAEnvK23'LTR were constructed by using pCMV-HA (Clontech) as a backbone, which contains an immediate-early promoter of human cytomegalovirus, the simian virus 40 (SV40) late 16S mRNA intron, and the SV40 polyadenylation signal. *REBK1*, *REBK2*, BERV-K1 *env*, and BERV-K2 *env* cDNAs with or without the respective 3' LTR were each ligated into pCMV-HA, which was digested with EcoRI and NotI. Schematic diagrams of the constructs are shown in Fig. 5A. Nucleotide structures of all constructs were verified by nucleotide sequencing analyses.

RT-PCR. Total RNA was extracted from frozen placenta and BT-1 cells with an RNeasy RNA isolation kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Reverse transcription (RT) was performed with an oligo(dT)₂₀ primer using a first-strand cDNA synthesis kit (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. PCR was performed with primers K1S (5'-GGATGCAATTTGGATCCAGAC-3') and K1R (5'-CCTT TGCATATTAGGCTCTCCG-3') for BERV-K1 *env*; K2S (5'-TCCACAGGA CGCAGATTCTCC-3') and K2R (5'-CCTTTGCGTATGCCGAGCCTCCT-3') for BERV-K2 *env*; BAS (5'-TACAATGAGCTGCGTGTGG-3') and BAR (5'-TAGCTCTTCCAGGGAGGA-3') for β -actin; 1S (5'-GATCCGGTACTAG

AGGAACTGAAAAAC-3') and 1R (5'-CACTGCATTCTAGTTGTGGTTTGTCC-3') for transcripts from pHAREBK1, pHAREBK2, pHAEnvK1, and pHAEnvK2; and 1S and K1LTR (5'-ATGTGGAACTTGGCAGCAAGC-3') or K2LTR (5'-AGTAAAGTCAGAGGTGAGAGCAC-3') for those from pHAEnvK13'LTR and pHAEnvK23'LTR, respectively. PCR conditions were as follows: the reaction mixture (50 μl) consisted of 2 μl cDNA template, 5 μl of $10\times$ buffer containing 20 mM MgCl_2 (Ex *Taq* buffer; TaKaRa, Shiga, Otsu, Japan), 0.25 μl *Taq* polymerase (Ex *Taq*; Takara), 4 μl of 2.5 mM deoxynucleotide triphosphates, 1 μl of each primer (100 pmol/ μl), and 36.75 μl distilled water. The amplification conditions were 94°C for 2 min, followed by 35 cycles of amplification consisting of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 2 min 30 s, and then a final extension step at 72°C for 7 min. PCR was carried out with 200- μl thin-walled PCR tubes using a Robocycler gradient 96 (Stratagene, La Jolla, CA).

Quantitative real-time RT-PCR. Total RNA was extracted from frozen tissues with TRIzol reagent (Invitrogen) according to the manufacturer's instructions. One microgram of total RNA, which was treated with DNase I (TaKaRa), was reverse transcribed into cDNA with a random primer (Toyobo, Osaka, Japan) and Superscript III reverse transcriptase (Invitrogen). Real-time RT-PCR was performed by using the TaqMan universal PCR master mix (Applied Biosystems) according to the manufacturer's instruction. PCR and the resulting increase in reporter fluorescent dye emission were monitored in real time by using an ABI Prism 7000 real-time PCR system (Applied Biosystems). The primer pairs and TaqMan probes were designed by the Primer Express program (Applied Biosystems). The primer pairs and probes used were as follows: forward primer 5'-GGAAATCACCGGATGTCCT-3', reverse primer 5'-GGAGAGGAGGCGCTTACCTG-3', and probe 5'-6-carboxyfluorescein (FAM)-CTGCAGGGCTCTCCAGTCTCTCCG-6-carboxymethyltetra rhodamine (TAMRA)-3' for BERV-K1 *env*; forward primer 5'-AAAGGAGGTC AGGCGCCTG-3', reverse primer 5'-TGGGGGAGGAGGCGCTTACCT-3', and probe 5'-FAM-CTCCCTGAGAAGATGTTTATTGCT-TAMRA-3' for BERV-K2 *env*; and forward primer 5'-AAGGCCATACCATTCTCA-3', reverse primer 5'-CCACCATACTCAGCACCAGCAT-3' (26), and probe 5'-FAM-A GCGAGATCTGCCAACATCAAGTGG-TAMRA-3' for the glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene. Primer pairs designed for BERV-K1 and -K2 were specific for the detection of *env* mRNAs but not *REBK* mRNAs. The thermal cycling conditions included 1 cycle at 50°C for 2 min, 1 cycle at 95°C for 10 min, and 40 cycles at 95°C for 15 s and 60°C for 1 min. To quantify the mRNA concentrations, standard curves were generated for each gene by the serial dilution of the plasmids containing their cDNAs. The expression ratio of each gene was normalized relative to the abundance of a validated endogenous control *GAPDH* mRNA to adjust for variations in the RT-PCRs. Quantitation was performed in duplicate for each tissue sample, and all values are presented as means \pm standard errors of the means (SEM).

Cloning and nucleotide sequences. Total RNA from the placenta was subjected to RT and PCR procedures as described above. PCR conditions were as follows: the reaction mixture (50 μl) consisted of 2 μl cDNA template, 5 μl of $10\times$ buffer containing 20 mM MgCl_2 (PfuUltra II buffer; Stratagene), 1 μl *Pfu* polymerase (PfuUltra II fusion HS DNA polymerase; Stratagene), 0.5 μl of 25 mM deoxynucleotide triphosphates, 1 μl of each primer (K1S/K1R or K2S/K2R; 100 pmol/ μl), and 39.5 μl distilled water. The amplification conditions were 95°C for 2 min, followed by 35 cycles of amplification consisting of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 1 min and then a final extension step at 72°C for 5 min. PCR was carried out with 200- μl thin-walled tubes using a Robocycler gradient 96 (Stratagene). The PCR products were analyzed by electrophoresis in a 1% agarose gel, and bands of the expected sizes were extracted from the gel with a PCR gel extraction kit (Qiagen). The extracted cDNA was cloned into the pCR-Blunt II-TOPO vector (Invitrogen), of which nucleotides were sequenced through the Shimadzu Corporation (Kyoto, Japan).

Fluorescence microscopy. HEK293T and PK-15 cells seeded into six-well culture plates were cultured for 24 h and then transfected with 1 μg pREBK1GFP, pREBK2GFP, and pAcGFP1-N1 (empty) using 3 μl FuGENE 6 transfection reagent (Roche Diagnostics, Indianapolis, IN). Forty-eight hours after transfection, cells were examined by using fluorescence microscopy. Prior to the inoculation of pseudotype viruses, HEK293T, PK-15, and MDBK cells were subcultured in six-well culture plates for 24 h. They were then inoculated with retroviral vectors, which were prepared by using the ecotropic MLV-based packaging cell line, Plat-E cells, a vesicular stomatitis virus G protein (VSV-G) expression plasmid (pCAGVSV-G), and each retroviral vector plasmid. The pseudotype viruses consist of an MLV core and a VSV-G envelope protein and contain the GFP, REBK1GFP, or REBK2GFP gene, respectively. At 48 h postinoculation, the cells were subjected to fluorescence microscopy analysis. Hoechst 33342 dye

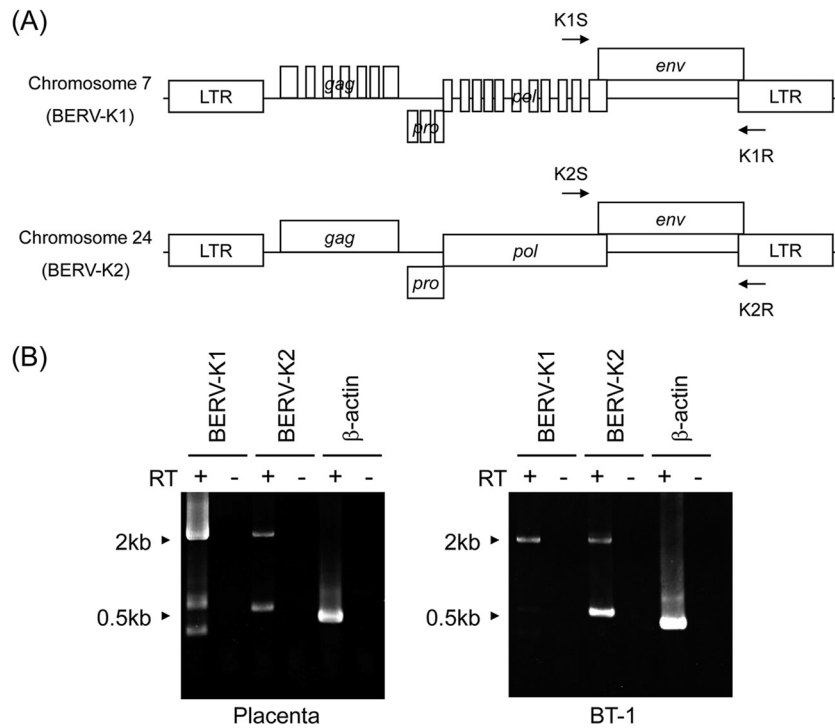


FIG. 1. Genomic structures of BERV-K1 and -K2 identified by *in silico* analyses. (A) Schematic representation of BERV-K1 and -K2 proviruses on bovine chromosomes 7 and 24, respectively. Arrows indicate primers used for RT-PCR analysis in B. (B) Expression of BERV-K1 and -K2 mRNAs in bovine placenta and BT-1 cells. Total RNA was extracted from the placenta and BT-1 cells and then subjected to RT-PCR analysis using the primers indicated in A. β -Actin was also amplified as an internal control.

(Nacalai Tesque, Kyoto, Japan) was added into each culture medium to stain cell nuclei before examination.

Dual-luciferase reporter assay. Firefly luciferase sequences were inserted between the splice donor (SD) and splice acceptor (SA) of start codon-deleted BERV-K *env* sequences with or without the 3' LTR, and these constructs were named pCMVK1Luc (or pCMVK2Luc) and pCMVK1Luc3'LTR (or pCMVK2Luc3'LTR), respectively (see Fig. 6A). HEK293T cells (1.5×10^5 cells/well) and PK-15 cells (3.0×10^4 cells/well) were subcultured on 48-well culture plates 24 h prior to transfection. pCMVKLuc or pCMVKLuc3'LTR constructs (50 and 250 ng/well for HEK293T and PK-15 cells, respectively) were then cotransfected with the pRL-TK vector (5 and 25 ng/well for HEK293T and PK-15 cells, respectively) by using FuGENE6 (Roche Diagnostics, Basel, Switzerland). To study the REBK functions, pREBK1GFP or pREBK2GFP was also cotransfected with the reporter plasmids at each amount, as described in Fig. 7, and pAcGFP1-N1 was also cotransfected at each amount to adjust the amount of total transfected DNA. Transfected cells were harvested after 48 h of incubation, and luciferase activities were measured with a Lumat LB9507 instrument (Berthold Technologies, Calmbacher, Bad Wildbad, Germany). Luciferase activity was measured by using a dual-luciferase reporter assay system (Promega, Madison, WI) according to the manufacturer's instructions. Assays were performed in triplicate and repeated as three independent experiments, and values of the luciferase activities are shown as the means \pm standard errors. Significant differences were assessed by a Student's *t* test for Fig. 6 and by the Tukey-Kramer multiple-comparisons procedure for Fig. 7, and the differences were considered significant at a *P* value of <0.05 .

Immunoblot analysis. HEK293T cells (4×10^5 cells/well) seeded into six-well culture plates were cultured for 24 h and then transfected with 1 μ g pHAREBK1, pHAREBK2, pHAEnvK1, pHAEnvK2, pHAEnvK13'LTR, or pHAEnvK23'LTR using 3 μ l FuGENE 6 transfection reagent (Roche Diagnostics). Forty-eight hours after transfection, the cells were harvested and subjected to RT-PCR (see above) and immunoblot analysis. Cells were lysed with 50 μ l IPB lysis buffer (20 mM Tris-HCl [pH 7.5], 150 mM NaCl, 1% Triton-X, 5 mg/ml sodium deoxycholate, 0.05% sodium dodecyl sulfate [SDS], 1 mM phenylmethylsulfonyl fluoride). Lysates were incubated at 4°C for 15 min and then centrifuged at $10,000 \times g$ for 10 min. Supernatants of cell lysates (10 μ l) were mixed with 10 μ l of 2 \times SDS sample buffer (Bio-Rad Laboratories, Richmond, CA)

containing 5% β -mercaptoethanol. They were heated at 99°C for 5 min and then subjected to SDS-polyacrylamide gel electrophoresis on a 12.5% gel. After the fractionated proteins were transferred onto polyvinylidene difluoride membranes using a Trans-Blot SD cell (Bio-Rad Laboratories), membranes were incubated in Tris-buffered saline (TBS) (pH 7.4) and 5% skim milk at room temperature for 1 h. Immunostaining was performed with TBS containing 0.1% Tween 20 (TBST). The blots were probed with an anti-HA antibody (Clontech Laboratories) at 37°C for 1 h and then washed 6 times with TBST. The blots were incubated with a horseradish peroxidase (HRP)-conjugated mouse anti-rabbit IgG antibody (Pierce, Rockford, IL) at 37°C for 1 h and then washed 6 times with TBST. Finally, the probed proteins were detected by using a Super Signal West Femto maximum-sensitivity system (Pierce).

Nucleotide sequence accession numbers. The BERV-K1 and -K2 *env* sequences were deposited in GenBank and assigned accession numbers AB587259 and AB587260, respectively.

RESULTS

Identification of endogenous betaretroviruses in the bovine genome. To identify BERV proviruses containing a full-length *env* ORF in the genome, a BLAST search was conducted on NCBI bovine genome resources using TM sequences of known ERVs, including PERV, enJSRV, HERV-W, syncytin-A, and syncytin-B. The numbers of hits were 84, 160, 165, 161, and 157 for PERV, enJSRV, HERV-W, syncytin-A, and syncytin-B, respectively. By repeating the BLAST search using positive sequences, we successfully found two proviruses containing the entire *env* ORF on chromosomes 7 and 24 (Fig. 1A). On the basis of the nucleotide sequences for PBS, tRNA^{Lys} was used as a primer for RT (data not shown); therefore, we named these elements BERV-K1 and -K2, respectively. Although only the *env* ORF was found in BERV-K1, additional ORFs corre-

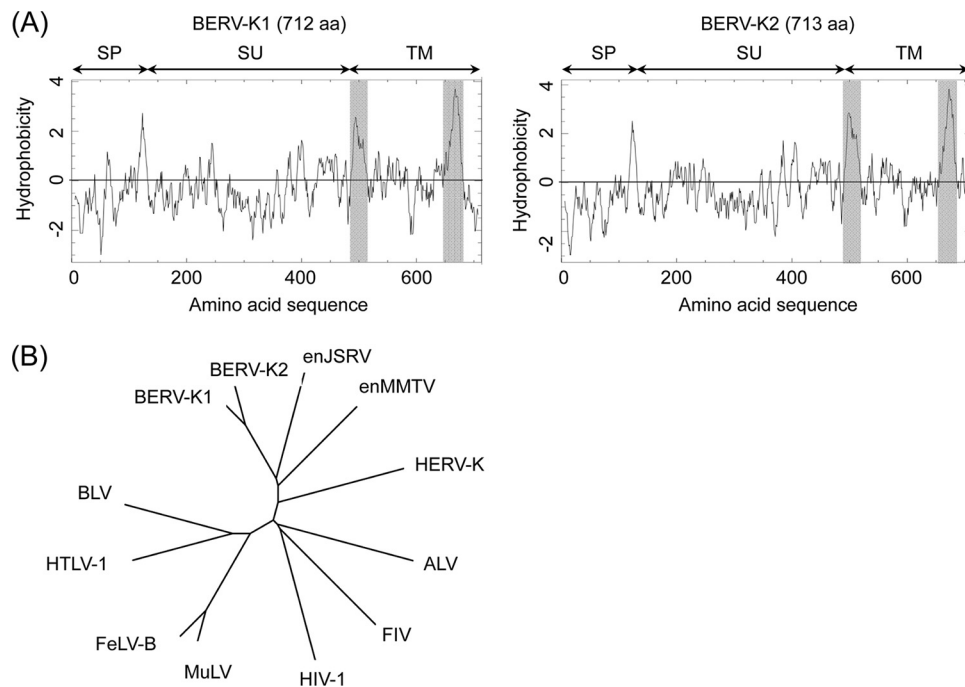


FIG. 2. Hydrophobicity profiles and phylogenetic analysis of Env proteins of BERV-K1 and -K2. (A) Hydrophobicity profiles of BERV-K1 and -K2 Env proteins. Gray shading indicates the fusion peptide and the transmembrane domain. (B) Phylogenetic analyses of the BERV-K1 and -K2 *env* genes. The phylogenetic tree of the entire TM-encoding region of *env* was constructed by the neighbor-joining method. enJSRV, endogenous Jaagsiekte sheep retrovirus; enMMTV, endogenous mouse mammary tumor virus; HERV-K, human endogenous retrovirus type K; ALV, avian leukemia virus; FIV, feline immunodeficiency virus; HIV-1, human immunodeficiency virus type 1; MuLV, murine leukemia virus; FeLV-B, feline leukemia virus subgroup B; HTLV-1, human T-cell leukemia virus type 1; BLV, bovine leukemia virus.

sponding to *gag*, *pro*, and *pol* were found in BERV-K2 (Fig. 1A). These genes possessed characteristic domains and motifs in retroviral proteins; *gag* encodes 725 amino acids (aa), containing p10 and p24 domains and two Cys Cys His Cys (CCHC)-type zinc finger motifs. *pro* encodes 203 aa, containing dUTPase, pepsin-like aspartate protease, and a G-patch domain. *pol* encodes 885 aa, containing reverse transcriptase (RTase), the RTase thumb, RNase H, the integrase zinc binding domain, the integrase core domain, and integrase (sequence data not shown).

Cloning and characterization of BERV-K1 and -K2 *env* genes. We found the expression of BERV-K1 and -K2 *env* mRNAs in bovine placenta and bovine trophoblast BT-1 cells (32) by RT-PCR using specific primers (Fig. 1B), and the amplicons were sequenced. The BERV-K1 and -K2 *env* genes encode 712 and 713 aa, respectively. Analysis of BERV-K1 and -K2 Env amino acid sequences revealed the presence of all characteristic domains and motifs of retroviral Env, including a signal peptide (SP), a surface (SU) Env protein, a TM Env protein (Fig. 2A), a hydrophobic fusion peptide, a membrane-spanning domain, a cleavage site by furin (R-X-K/R-R) between the SU and TM proteins, and two disulfide motifs (CXXC and CX7C) involved in the interaction between the SU and TM proteins (data not shown). These data indicate that the BERV-K1 and -K2 Env proteins can form a conformation typical of retroviral Env. Phylogenetic analysis revealed that both BERV-K1 and -K2 were closely related to enJSRV, belonging to genus *Betaretrovirus* (Fig. 2B).

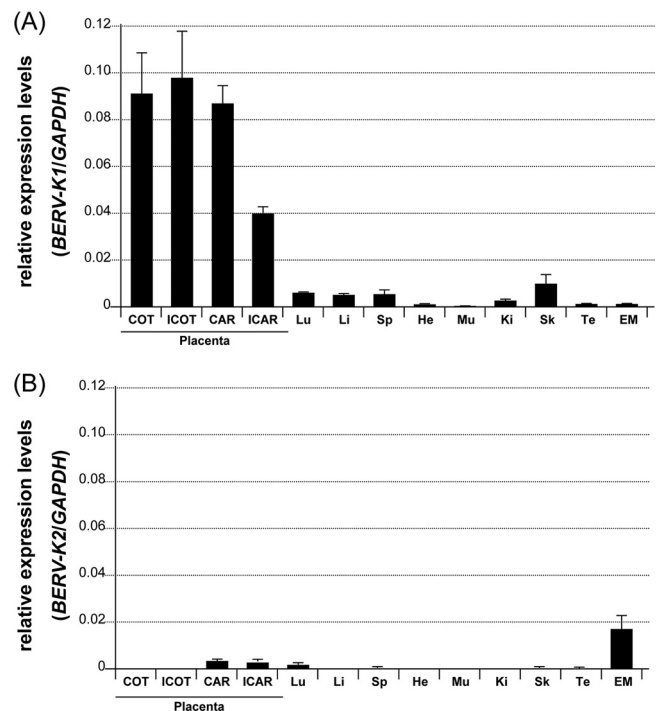


FIG. 3. Expression of BERV-K1 and -K2 *env* mRNAs in normal bovine tissues. Total RNA was extracted from tissues of Holstein cows. Abbreviations: COT, cotyledon; ICOT, intercotyledon; CAR, caruncular; ICAR, intercaruncular; Lu, lung; Li, liver; Sp, spleen; He, heart; Mu, muscle; Ki, kidney; Sk, skin; Te, testis; EM, endometrium. Each value was normalized to the amount of *GAPDH* mRNA and is exhibited as means \pm standard errors of data from three or six samples in duplicate experiments.

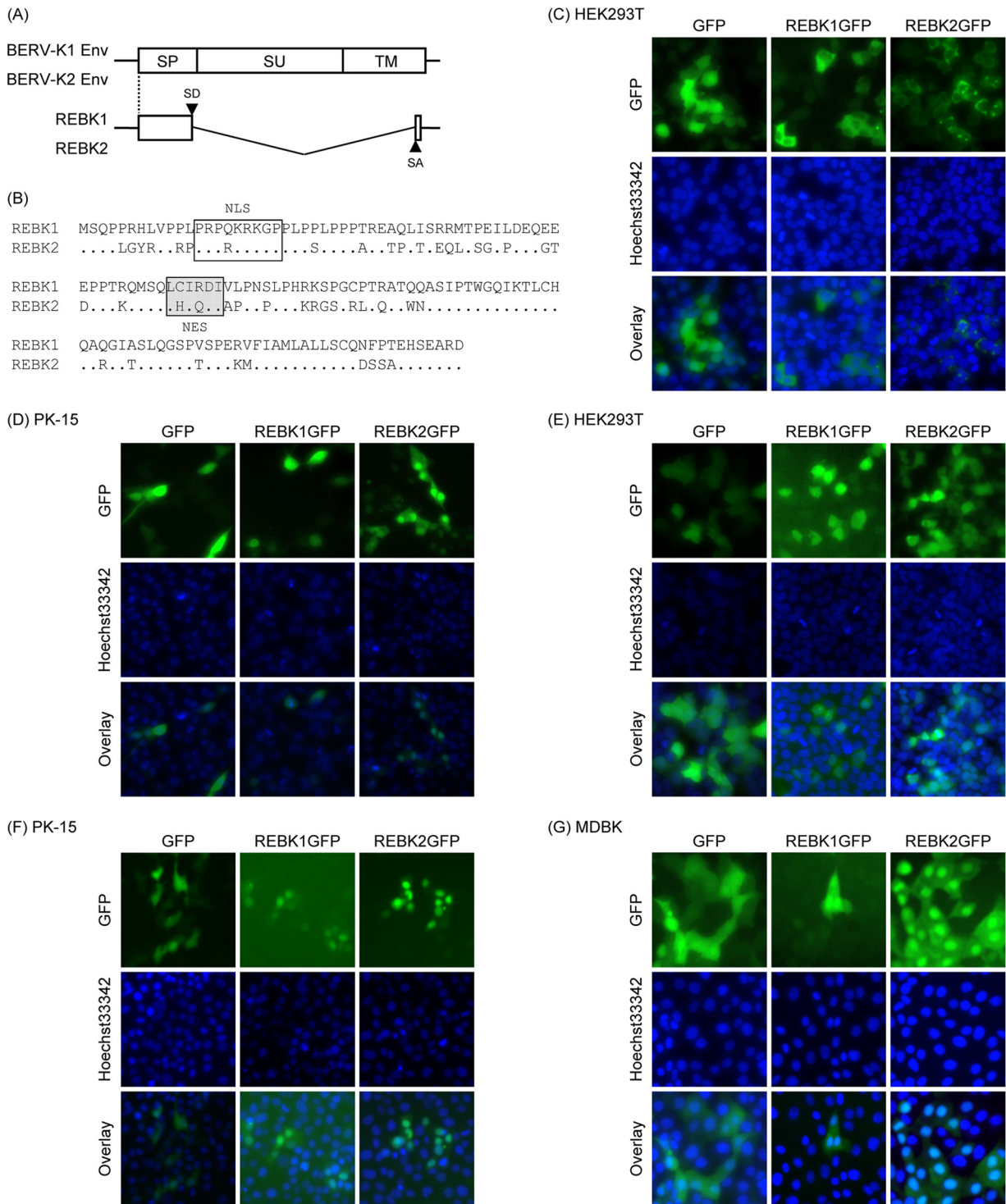


FIG. 4. Identification of the *REBK1* and *REBK2* proteins. (A) Schematic representation of *REBK1* and *REBK2* transcripts. Start codons for *REBK1* and *REBK2* are shared with those of *env*. (B) Sequence alignment of the *REBK1* and *REBK2* proteins. Dots indicate identical amino acids. The box and the shaded box indicate NLS and NES motifs, respectively. (C to G) Subcellular localization of the *REBK1* and *REBK2* proteins. *REBK1* and *REBK2* fused with GFP (*REBK1GFP* and *REBK2GFP*, respectively) or only GFP (empty vector) was expressed in HEK293T (C and E), PK-15 (D and F), and MDBK (G) cells by transfection with expression plasmids (C and D) and infection with retroviral vectors (E, F, and G). Forty-eight hours after transfection and infection, the cells were stained with Hoechst 33342 dye and subjected to fluorescence microscopy analysis.

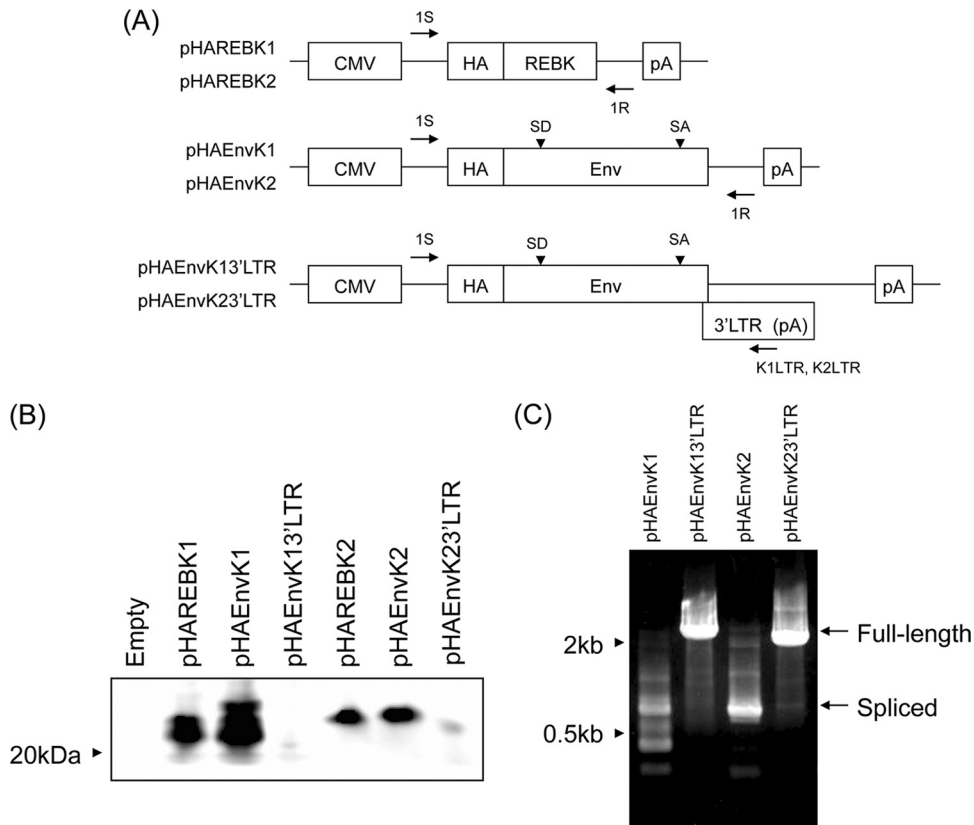


FIG. 5. Splicing of BERV-K *env* mRNA depending on the 3' LTR sequence. (A) Schematic representation of expression plasmids. Primers used for RT-PCR in C are indicated by arrows. (B) Immunoblot analysis for the expression of REBK1 and REBK2. HEK293T cells were transfected with the indicated expression plasmids. Forty-eight hours after transfection, cell lysates were extracted and subjected to immunoblot analysis. HA-tagged REBK1 and REBK2 were probed with a rabbit anti-HA IgG antibody and detected by an HRP-labeled mouse anti-rabbit IgG antibody. (C) Detection of full-length and spliced transcripts. Total RNAs were extracted from cells transfected with pHAEnvK1, pHAEnvK2, pHAEnvK13'LTR, and pHAEnvK23'LTR and subjected to RT-PCR analysis. Full-length *env* and singly spliced transcripts are indicated by arrows.

Expression of BERV-K1 and -K2 *env* mRNAs in bovine tissues. We examined the expression levels of *env* mRNAs derived from loci of BERV-K1 and -K2 in various bovine tissues by real-time RT-PCR. Whereas BERV-K1 *env* mRNAs were preferentially expressed in placental tissues, especially in the placentome (COT and CAR) and ICOT, BERV-K2 *env* mRNAs were marginally expressed in various tissues, with the exception of the endometrium (Fig. 3).

Identification of a small ORF generated by a double-splicing event of the *env* mRNA. As shown in Fig. 1B, we found additional transcripts shorter than the full-length *env* mRNAs in the placenta by RT-PCR. Nucleotide sequencing analyses revealed that these transcripts were expressed by an alternate internal splicing event in *env* mRNAs (Fig. 4A) (GenBank accession numbers AB587261 and AB587262). The ORFs started with the same start codon of *env* and encoded 140 aa consisting of an N-terminal region essentially identical to SP (1 to 129 amino acids in Env) and an additional C-terminal 11 aa (Fig. 4A and B). Putative NLS and NES motifs were also found in the predicted amino acid sequences (Fig. 4B). Because these motifs were similar to those found in other retroviral regulatory proteins, we designated the predicted proteins the RNA export protein-like proteins of BERV-K (REBK1 and

REBK2). When the REBK1 and REBK2 proteins fused with GFP (REBK1GFP and REBK2GFP, respectively) were expressed by the transfection of expression plasmids, the proteins were localized in the nuclei of porcine cells (PK-15 cells) (Fig. 4D), whereas they were localized diffusely in human HEK293T cells (Fig. 4C). Because these expression plasmids contain an SV40 *ori*, we suspected that the proteins might have been overexpressed and thus localized mainly in the cytoplasm in HEK293T cells. Therefore, we introduced the REBK genes fused with GFP without the SV40 *ori* sequence by using an MLV-based retroviral vector. When these proteins were expressed through retroviral vectors, both the REBK1GFP and REBK2GFP proteins were localized mainly in the nuclei in both HEK293T and PK-15 cells (Fig. 4E and F). Furthermore, we also confirmed that both the REBK1GFP and REBK2GFP proteins were localized mainly in the nuclei in MDBK cells, which are a bovine kidney cell line (Fig. 4G).

Existence of a regulatory element in the 3' LTR responsible for the splicing of viral RNA. To express BERV-K Env proteins, we constructed expression plasmids of BERV-K Env fused with an HA tag at the N terminus, termed pHAEnvK1 and pHAEnvK2 (Fig. 5A). When we transfected these plasmids into HEK293T cells, we detected approximately 20-kDa

proteins but not the SU Env (approximately 60 kDa) or the precursor Env (approximately 80 kDa) proteins. To determine whether there is a regulatory element in the 3' LTR of the BERV-K genome responsible for splicing *env* mRNA, we also constructed expression plasmids of HA-tagged BERV-K Env proteins with the 3' LTR, termed pHAEnvK13'LTR and pHAEnvK23'LTR, and expression plasmids of REBK proteins fused with an HA tag (Fig. 5A). By immunoblot analysis, the expression of the HA-tagged REBK protein (approximately 20kDa) was detected in HEK293T cells transfected with pHAREBK and pHAEnvK plasmids but not in those transfected with pHAEnvK3'LTR plasmids (Fig. 5B). Although spliced mRNAs were detected only in cells transfected with pHAEnvK plasmids, full-length *env* mRNA was detected in cells transfected with pHAEnvK3'LTR plasmids (Fig. 5C). This result indicates that the 3' LTRs of the BERV-K genomes inhibit the second splicing that occurred within the *env* regions. We could not detect the Env proteins by transfection with pHAEnvK13'LTR and pHAEnvK23'LTR in HEK293T cells by immunoblot analysis (Fig. 5B). The exact reason for this observation is unknown at present; however, an additional factor(s) may be required for the efficient expression of the Env proteins.

Double splicing was suppressed by 3' LTR sequences. To confirm the splicing regulation shown in Fig. 5, reporter plasmids, termed pCMVK1Luc, pCMVK1Luc3'LTR, pCMVK2Luc, and pCMVK2Luc3'LTR, were constructed (Fig. 6A) and then subjected to a dual-luciferase reporter assay (Fig. 6B). These reporter plasmids do not express REBK proteins because the N termini of *REBK* genes, including the initiation codons for the REBK proteins, are deleted. The relative luciferase activities should parallel the levels of unspliced mRNAs transcribed from the plasmids. Because the transfection efficiency was very low in MDBK cells, we used HEK293T and PK-15 cells in the following studies of the functions of the 3' LTR and REBK proteins. pCMVK1Luc3'LTR exhibited a 1.9-fold-higher luciferase activity than did pCMVK1Luc in HEK293T cells ($P < 0.05$); however, there was no difference in PK-15 cells. In contrast, pCMVK2Luc3'LTR exhibited 19.7- and 4.7-fold-higher luciferase activities than did pCMVK2Luc in HEK293T and PK-15 cells, respectively ($P < 0.05$). These results indicate that there are viral constitutive transport element (CTE)-like elements (3, 31) in the 3' LTRs of both BERV-K1 and -K2.

Functional studies of REBKs. To investigate whether REBKs are needed to export unspliced mRNA and/or subsequent translation, pREBK1GFP plasmids were cotransfected with pCMVK1Luc plasmids or pCMVK1Luc3'LTR plasmids, and the dual-luciferase reporter assay was then conducted (Fig. 7). pREBK1GFP augmented luciferase activity expressed from pCMVK1Luc in a dose-dependent manner in both HEK293T and PK-15 cells. On the other hand, pREBK1GFP augmented luciferase activity expressed from pCMVK1Luc3'LTR only in HEK293T cells. In contrast, pREBK2GFP did not affect luciferase activities expressed from pCMVK2Luc in either HEK293T or PK-15 cells but augmented the activity expressed from pCMVK2Luc3'LTR in HEK293T cells. It should be noted that there is no dose dependency of the effect.

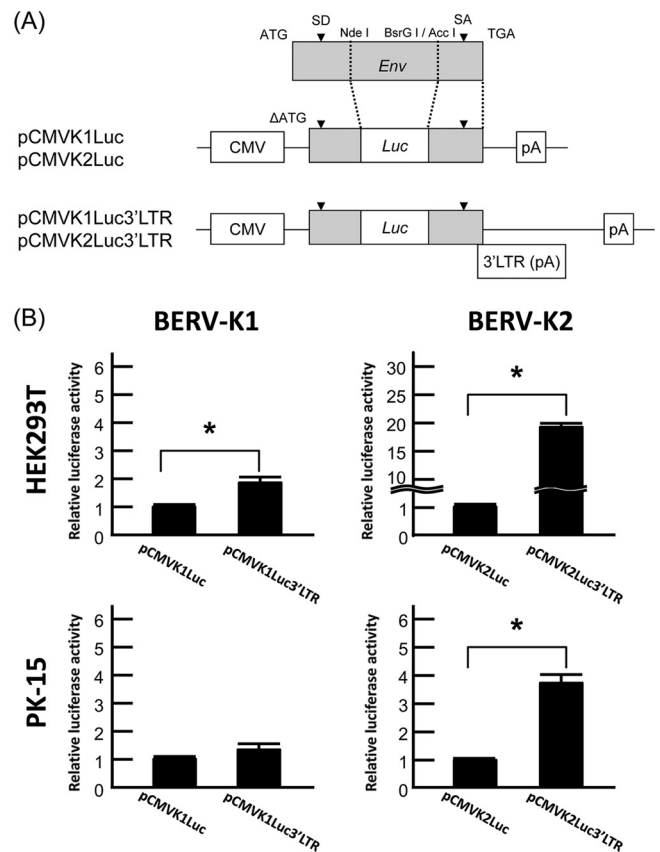


FIG. 6. Splicing of BERV-K *env* mRNAs is suppressed by the 3' LTR sequence. (A) Schematic diagrams of reporter plasmids. Plasmids contain partial sequences of BERV-K1 or BERV-K2 *env* without translation starting sequences (Δ ATG), and the luciferase gene is inserted between the SD and SA sites of *env*. pCMVK1Luc plasmids are Δ 3' LTR mutants of pCMVK1Luc3'LTR plasmids. CMV, cytomegalovirus. (B) The relative luciferase activity of each reporter plasmid was measured in HEK293T and PK-15 cells. Values are the means \pm standard errors of data from three independent experiments in triplicate. Asterisks indicate significant differences ($P < 0.05$).

DISCUSSION

This is the first report of the identification of bovine ERVs containing intact ORFs. BERV-K1 and -K2 identified in this study were classified into the genus *Betaretrovirus* (Fig. 2). Estimations from the identical residues shared between the 5' and 3' LTRs suggested that BERV-K1 integrated into the bovine genome much earlier than did BERV-K2 (16). Consistent with this estimation, *gag*, *pro*, *pol*, and *env* ORFs with the characteristic domains and motifs were maintained in BERV-K2, whereas only the *env* ORF was maintained in BERV-K1 (Fig. 1). These data suggest that viral proteins of BERV-K1 and -K2 may still have unique functions and that the BERV-K2 provirus on chromosome 24 may have the ability to produce viral particles.

We found the expression of BERV-K1 and -K2 *env* mRNAs in the placenta and a trophoblast cell line by RT-PCR (Fig. 1B). Real-time RT-PCR revealed that the BERV-K1 *env* mRNA was preferentially expressed in the placenta; however, the levels of BERV-K2 *env* expression were much lower than those of BERV-K1 *env* in most tissues (Fig. 3). Although we do

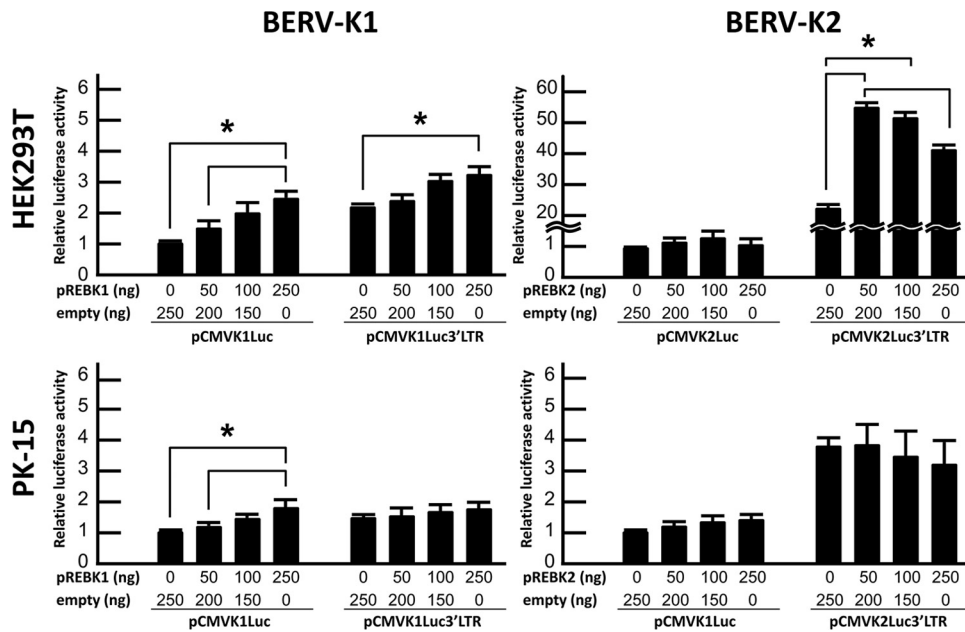


FIG. 7. Effects of REBKs on the splicing of BERV-K *env* mRNAs. Each graph indicates the relative luciferase activity of each reporter plasmid cotransfected with variable amounts of pREBKGFp plasmids (abbreviated pREBK1 and pREBK2) and pAcGFP1-N1 (noted as empty) in HEK293T and PK-15 cells. Values are the means \pm standard errors of data from three independent experiments in triplicate. Asterisks indicate significant differences ($P < 0.05$).

not know the precise mechanisms by which BERV-K2 expression was suppressed in bovine tissues, contrary to the high BERV-K1 expression levels in the bovine placenta, there should be some epigenetic regulatory systems like those of other ERVs (15, 18, 21). Interestingly, the functional roles of Env proteins in placental morphogenesis may differ in mammalian species which have different morphogenetic patterns of placenta. Humans, mice, and rabbits have a hemochorial placenta, whereas sheep and cattle have a syndesmochorial (which is essentially an epitheliochorial) placenta. HERV-W (syncytin-1) and HERV-FRD (syncytin-2) Env proteins were found previously to induce the fusion of trophoblast cells, followed by the formation of syncytiotrophoblasts (2, 23), similarly to syncytin-A and syncytin-B in mice (8) and syncytin-Ory I in rabbits (11). On the other hand, enJSRV Env proteins have been shown to regulate the growth and proliferation of trophoblast cells in peri-implantation conceptuses (7, 28). The physiological activities of Env proteins (e.g., fusogenicity and signal transduction) could be associated with the morphological evolution of placental structures in mammals. Extensive experiments are required to further understand the roles of the Env proteins of BERV-K1 and -K2 in bovine placental morphogenesis.

The strategy for exporting unspliced viral RNA is different for betaretroviruses. HERV-K and mouse mammary tumor virus encode the *trans*-acting regulatory proteins Rec and Rem, respectively (13, 19, 20, 22). On the other hand, Mason-Pfizer monkey virus and type D simian retrovirus contain the *cis*-acting CTE at the 3' end of viral genomic RNA, since they do not have *trans*-acting regulatory proteins (3, 31). Moreover, JSRV contains both a CTE-like element and sequences that respond to Rej, which facilitates the translation of unspliced viral mRNA predominantly and the nuclear export of un-

spliced viral RNA modestly (4, 12, 27). In this study, we also found transcripts, named *REBK*, expressed by alternate splicing events in the *env* mRNA other than full-length *env* mRNAs in the placenta (Fig. 1B and data not shown). The REBK1 and REBK2 proteins have putative NLS and NES motifs (data not shown) and were localized in the nuclei (Fig. 4 and data not shown) in a manner similar to that observed for the Rec and Rem proteins.

In our preliminary study, we could not detect precursor and TM Env proteins of BERV-K1 and -K2 in HEK293T cells by transfection with expression plasmids containing the entire *env* ORF with a Myc-His tag in the C terminus, whereas SU Env proteins were successfully expressed by transfection with an expression plasmid containing the C-terminally tagged SU region (data not shown). The results shown in Fig. 5 suggest that there may be a responsive element between the C-terminal region of the *env* ORF and the downstream 3' LTR region necessary for the nuclear export of *env* mRNA and/or the efficient translation of Env in the cytoplasm. It is plausible that the failure of the expression of BERV-K Env tagged with Myc-His in the C terminus is caused by the loss of function of the responsible elements when tag sequences were introduced into the regions.

To elucidate the CTE-like elements and the functional interaction between the REBK-responsive elements and REBKs, we performed luciferase reporter assays (Fig. 6 and 7). The data shown in Fig. 6 indicate that the CTE-like elements are present in the 3' LTRs of both BERV-K1 and -K2. In addition, the CTE-like activity was prominent for BERV-K2 in both HEK293T and PK-15 cells but minimal and inactive for BERV-K1 in HEK293T and PK-15 cells, respectively. Both BERV-K1 and -K2 have REBK-responsive elements in the region encompassing *env* and the 3' LTR (Fig. 7). The 3' LTR

of BERV-K2 contains both a CTE-like element and a REBK-responsive element; however, the REBK-responsive element is functional only in HEK293T cells. On the other hand, BERV-K1 has a REBK1-responsive element with minimal REBK1 activities, but the responsive element may reside in the *env* region. Although we identified the CTE-like elements and REBK-responsive elements in BERV-K1 and -K2 using *env* expression plasmids, it is still unknown whether the singly spliced *env* mRNAs transcribed from the loci of BERV-K1 and -K2 are subject to the effects of REBK1 and REBK2 and/or CTE-like elements. For the other analogous betaretrovirus regulatory proteins, their function is to export unspliced viral RNA to the cytoplasm and/or enhance its translation (5). Usually, the singly spliced *env* mRNAs of other retroviruses can exit the nucleus via the pathway used by spliced cellular mRNAs (5). It is noteworthy that the more ancient BERV-K1, which has retained only the *env* ORF, has both attenuated CTE and REBK activities, while BERV-K2 has more-robust activities. This could reflect a role for the CTE and REBKs in unspliced mRNA transport/translation, which is still present in the more recently integrated BERV-K2, while they are not retained for BERV-K1, whose Env is being expressed via the cellular mRNA splicing/transport pathway. Further studies using the entire genomes of BERV-Ks are needed to elucidate these issues.

In summary, we identified two novel endogenous betaretroviruses, named BERV-K1 and -K2, in the bovine genome. Both BERV-K1 and -K2 have intact *env* ORFs, and the expression of Env proteins may be regulated by the CTE-like element and the regulatory proteins REBK1 and REBK2. Although we have not identified the Env proteins expressed in the bovine placenta, the data presented here imply that both BERV-K1 and -K2 express Env proteins *in vivo*, and the proteins may have physiological functions, especially in the placenta.

ACKNOWLEDGMENTS

We thank Peter Gee (Institute for Virus Research, Kyoto University) for his generous help in the preparation of the manuscript.

This study was supported by grants-in-aid from the Bio-Oriented Technology Research Advancement Institution and the Agri-Bio Project of the Ministry of Agriculture, Forestry, and Fisheries of Japan. Y.N. is supported by the Japan Society for the Promotion of Science for Japanese Junior Scientists.

REFERENCES

- Akiyoshi, D. E., et al. 1998. Identification of a full-length cDNA for an endogenous retrovirus of miniature swine. *J. Virol.* **72**:4503–4507.
- Blaise, S., N. de Parseval, L. Benit, and T. Heidmann. 2003. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc. Natl. Acad. Sci. U. S. A.* **100**:13013–13018.
- Bray, M., et al. 1994. A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. *Proc. Natl. Acad. Sci. U. S. A.* **91**:1256–1260.
- Caporale, M., et al. 2009. The signal peptide of a simple retrovirus envelope functions as a posttranscriptional regulator of viral gene expression. *J. Virol.* **83**:4591–4604.
- Cullen, B. R. 2003. Nuclear mRNA export: insights from virology. *Trends Biochem. Sci.* **28**:419–424.
- DeMartini, J. C., J. O. Carlson, C. Leroux, T. Spencer, and M. Palmarini. 2003. Endogenous retroviruses related to Jaagsiekte sheep retrovirus. *Curr. Top. Microbiol. Immunol.* **275**:117–137.
- Dunlap, K. A., et al. 2006. Endogenous retroviruses regulate periimplantation placental growth and differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **103**:14390–14395.
- Dupressoir, A., et al. 2005. Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muriidae. *Proc. Natl. Acad. Sci. U. S. A.* **102**:725–730.
- Dupressoir, A., et al. 2009. Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc. Natl. Acad. Sci. U. S. A.* **106**:12127–12132.
- Hecht, S. J., K. E. Stedman, J. O. Carlson, and J. C. DeMartini. 1996. Distribution of endogenous type B and type D sheep retrovirus sequences in ungulates and other mammals. *Proc. Natl. Acad. Sci. U. S. A.* **93**:3297–3302.
- Heidmann, O., C. Vernochet, A. Dupressoir, and T. Heidmann. 2009. Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new “syncytin” in a third order of mammals. *Retrovirology* **6**:107.
- Hofacre, A., T. Nitta, and H. Fan. 2009. Jaagsiekte sheep retrovirus encodes a regulatory factor, Rej, required for synthesis of Gag protein. *J. Virol.* **83**:12483–12498.
- Indik, S., W. H. Gunzburg, B. Salmons, and F. Rouault. 2005. A novel, mouse mammary tumor virus encoded protein with Rev-like properties. *Virology* **337**:1–6.
- Lander, E. S., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–921.
- Lavie, L., M. Kitova, E. Maldener, E. Meese, and J. Mayer. 2005. CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2). *J. Virol.* **79**:876–883.
- Lebedev, Y. B., et al. 2000. Differences in HERV-K LTR insertions in orthologous loci of humans and great apes. *Gene* **247**:265–277.
- Le Tissier, P., J. P. Stoye, Y. Takeuchi, C. Patience, and R. A. Weiss. 1997. Two sets of human-tropic pig retrovirus. *Nature* **389**:681–682.
- Liang, C. Y., et al. 2010. GCM1 regulation of the expression of syncytin 2 and its cognate receptor MFSD2A in human placenta. *Biol. Reprod.* **83**:387–395.
- Lower, R., R. R. Tonjes, C. Korbmayer, R. Kurth, and J. Lower. 1995. Identification of a Rev-related protein by analysis of spliced transcripts of the human endogenous retroviruses HTDV/HERV-K. *J. Virol.* **69**:141–149.
- Magin, C., R. Lower, and J. Lower. 1999. cORF and RcRE, the Rev/Rex and RRE/RxRE homologues of the human endogenous retrovirus family HTDV/HERV-K. *J. Virol.* **73**:9496–9507.
- Matoušková, M., J. Blazková, P. Pajer, A. Pavlíček, and J. Hejnar. 2006. CpG methylation suppresses transcriptional activity of human syncytin-1 in non-placental tissues. *Exp. Cell Res.* **312**:1011–1020.
- Mertz, J. A., M. S. Simper, M. M. Lozano, S. M. Payne, and J. P. Dudley. 2005. Mouse mammary tumor virus encodes a self-regulatory RNA export protein and is a complex retrovirus. *J. Virol.* **79**:14737–14747.
- Mi, S., et al. 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**:785–789.
- Morita, S., T. Kojima, and T. Kitamura. 2000. Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther.* **7**:1063–1066.
- Morozov, V. A., A. V. Morozov, and S. Lagaye. 2007. Endogenous JSRV-like proviruses in domestic cattle: analysis of sequences and transcripts. *Virology* **367**:59–70.
- Nakaya, Y., K. Kizaki, T. Takahashi, O. V. Patel, and K. Hashizume. 2009. The characterization of DNA methylation-mediated regulation of bovine placental lactogen and bovine prolactin-related protein-1 genes. *BMC Mol. Biol.* **10**:19.
- Nitta, T., A. Hofacre, S. Hull, and H. Fan. 2009. Identification and mutational analysis of a Rej response element in Jaagsiekte sheep retrovirus RNA. *J. Virol.* **83**:12499–12511.
- Palmarini, M., et al. 2001. Expression of endogenous betaretroviruses in the ovine uterus: effects of neonatal age, estrous cycle, pregnancy, and progesterone. *J. Virol.* **75**:11319–11327.
- Palmarini, M., M. Mura, and T. E. Spencer. 2004. Endogenous betaretroviruses of sheep: teaching new lessons in retroviral interference and adaptation. *J. Gen. Virol.* **85**:1–13.
- Patience, C., et al. 2001. Multiple groups of novel retroviral genomes in pigs and related species. *J. Virol.* **75**:2771–2775.
- Rizvi, T. A., R. D. Schmidt, K. A. Lew, and M. E. Keeling. 1996. Rev/RRE-independent Mason-Pfizer monkey virus constitutive transport element-dependent propagation of SIVmac239 vectors using a single round of replication assay. *Virology* **222**:457–463.
- Shimada, A., H. Nakano, T. Takahashi, K. Imai, and K. Hashizume. 2001. Isolation and characterization of a bovine blastocyst-derived trophoblastic cell line, BT-1: development of a culture system in the absence of feeder cell. *Placenta* **22**:652–662.
- Waterston, R. H., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**:520–562.
- Xiao, R., et al. 2008. Characterization of the bovine endogenous retrovirus beta3 genome. *Mol. Cells* **25**:142–147.
- Xiao, R., K. Park, Y. Oh, J. Kim, and C. Park. 2008. Structural characterization of the genome of BERV gamma4, the most abundant endogenous retrovirus family in cattle. *Mol. Cells* **26**:404–408.
- York, D. F., R. Vigne, D. W. Verwoerd, and G. Querat. 1992. Nucleotide sequence of the Jaagsiekte retrovirus, an exogenous and endogenous type D and B retrovirus of sheep and goats. *J. Virol.* **66**:4930–4939.