PNAS
www.pnas.org

Supplementary Information for

Convergent evolution of antiviral machinery derived from endogenous retrovirus truncated envelope genes in multiple species

Ariko Miyake, Minh Ha Ngo, Shelly Wulandari, Masayuki Shimojima, So Nakagawa, Junna Kawasaki, Kazuo Nishigaki*

*Corresponding author: Kazuo Nishigaki

Email: kaz@yamaguchi-u.ac.jp

Tel.: +81-83-933-5829

This PDF file includes:

Supplementary text

Figures S1 to S15

Tables S1 to S2

Dataset S1. SRA Run accession, genome file, and gene annotation file used for the mapping analysis.

Dataset S2. Sequence accession numbers used for phylogenetic analyses.

SI References

Supplementary Information Text

Supplementary Materials and Methods

Cell culture. Cells were cultured in high-glucose Dulbecco’s modified Eagle medium (FUJIFILM Wako Pure Chemical Corporation, Osaka, Japan) supplemented with 10% fetal bovine serum. Cells were incubated in a CO_2 incubator at 37°C. In this study, we used the following cell lines: HEK293T (human embryonic kidney transformed with SV40 large T antigen), MDTF (*Mus dunni* fibroblast tail), HepG2 (human), Cos7 (African green monkey), Vero (African green monkey), CRFK (cat), BHK-21(hamster), 104C1 (guinea pig), KwDM (dog), and MDBK (cattle) cells, as well as HEK293T cells persistently infected with feline leukemia virus (FeLV)-D (293T/FeLV-D), FeLV-B (293T/FeLV-B), endogenous retrovirus of domestic cats (ERV-DC)14TA (293T/ERV-DC14) that has a thymine-to-adenine mutation in the ERV-DC14 provirus LTR, and ERV-DC10 (293T/ERV-DC10) (1,2,3). Additionally, GPLac cells harboring a murine leukemia virus (MLV) Gag-Pol expression vector and a pMXs retroviral vector carrying a LacZ marker were cultured under the same conditions.

MDTF/feline copper transport protein 1 (feCTR1), MDTF/human copper transport protein 1 (huCTR1), MDTF/pFUΔss empty vector, MDTF/human copper transport protein 2 (huCTR2), and MDTF-pMSCVneo empty vector cell lines were established. Briefly, MDTF cells were transfected with expression vectors (pFUΔss-feCTR1 or pFUΔss empty vector as a control) using Lipofectamine®3000 reagent (Thermo Fisher Scientific, Waltham, MA, USA). Single cells were cultured in medium containing 200 µg/mL zeocin (InvivoGen, San Diego, CA, USA) in 96-well plates and single cells were isolated. PLAT-E packaging cells were transfected with expression vectors (pMSCVneo-huCTR1, pMSCVneo-huCTR2, or pMSCVneo empty vector) using TransIT®-293 reagent and supernatants were used to infect MDTF cells. Cells were cultured in medium containing 600 µg/mL neomycin. These cells were termed MDTF-feCTR1, MDTF-pFUΔss empty, MDTF-huCTR1, MDTF-huCTR2, and MDTF-pMSCVneo empty.

cDNA cloning and establishment of cell lines expressing CTR1. Feline RNA was extracted from the feline lymphoma FT-1 cell line using the RNeasy Mini Kit (Qiagen, Venlo, Netherlands). To obtain the feline CTR1 sequence, 5’ rapid amplification of cDNA ends was conducted using the 5’-Full RACE Core Set (Takara, Kusatsu, Japan) and the following primers: RT-primer (5’-CTGCTTCCCACTGC-3’, 5'-phosphorylated), S1-primer (5’-CAAGATAGCCCGAGAGAGCCTAC-3’), S2-primer (5’-GAGACACACAAAACTGTTGGGCA-3’), A1-primer (5’-AAAACACTGCCACGAAAGCTCC-3’), and A2-primer (5’- ACATCATCATACCGTCCATTCCCC-3’). We obtained approximately 1.0 kb of feline CTR1 sequence showing 100% homology with the predicted *Felis catus* solute carrier family 31 member 1 (SLC31A1) (accession number XM_023242310.1). Feline CTR1 cDNA was PCR-amplified using cDNA from the FT-1 cell line as a template with KOD One Master Mix (Toyobo, Osaka, Japan) and the primer pair Fe-652S (5’-ATTGGTGACGGGTTAAGGTTCCA-3’) and Fe-627R (5’-TAAGGCCACACCATGGAGTTGAACG-3’). The amplicon was cloned into the pCR4 Blunt-TOPO vector (Invitrogen, Carlsbad, CA, USA). To construct the feCTR1 expression plasmids, the primer pairs Fe-654S (5’-GAATTCGAATTCGCCGCCACCATGTACCCATACGATGTTCCAGATTACGCTCTCGAGATGGCTGACAACAGTACAACG-3’) and Fe-628R (5’-AGATCTAGATCTCTAATGGCAATGCTCTGTGATGTCCAC-3’), as well as Fe-654S and Fe-629R (5’-GCTAGCGCTAGCCTAATGGCAATGCTCTGTGATGTCCAC-3’), were used to amplify the feCTR1 ORF. The primers encoded *Eco*RI and *Bgl*II restriction sites for pMSCVneo (Takara) and *Eco*RI and *Nhe*I restriction sites for pFUΔss (InvivoGen). The feCTR1 ORF was fused with an N-terminal hemagglutinin (HA) tag. Human CTR1 was PCR-amplified from the HEK293 cDNA library using the primers hCTR1-cl-F1 (5’-TCTAGAATTCCTCGAGATGGATCATTCCCACCATATGG-3’) and hCTR1-cl-R1 (5’-TAACAGATCTGAATTCTCAATGGCAATGCTCTGTGATATCC-3’). The amplicon was digested with *Xho*I and *Bgl*II, and then replaced with the sequence of the pMSCVneo-feCTR1 plasmid, producing the pMSCVneo-huCTR1 plasmid. A DNA construct encoding N-terminally c-Myctagged huCTR2 was synthesized (GENEWIZ, Shinagawa, Japan) using the NCBI reference sequence (accession NP_001851.1). The cDNA was cloned into the pMSCVneo vector using *Eco*RI.

Quantitative RT-PCR of feline CTR1. Total RNA was extracted from the tissues of a specific pathogen-free cat (Kyoto-SPF1) and from the feline cell lines AH927, G355, CRFK, Fet-J, MCC, 3201, and MS4 using an RNAiso Plus kit (Takara), in accordance with the manufacturer's instructions. Thereafter, cDNA was synthesized using a PrimeScript II first-strand cDNA synthesis kit (Takara), in accordance with the manufacturer's instructions. Prior to reverse transcription, RNA samples were treated with recombinant DNase I (TaKaRa). cDNA was amplified with SYBR Premix Ex Taq II (Tli RNaseH Plus; Takara) in a CFX96 Touch real-time PCR detection system (Bio-Rad, Hercules, CA, USA). Feline CTR1 was amplified using primers Fe-657S (5’-GCCTACTGCGTAAGTCACAAG-3’) and Fe-632R (5’-TGCAGGAGGTGGGGAAAGCTC-3’). Feline peptidyl prolyl isomerase A (PPIA) was amplified using primers Fe-227S (5’-GTCAACCCCATCGTGTTTTT-3’) and Fe-204R (5’-CTGCTGTCTTGGGAACTTTGTC-3’).

Quantification of expression levels of truncated *env*, full-length *env*, and CTR1 in primate tissues using publicly available data. The expression levels of truncated *env*, full-length *env*, and CTR1 were analyzed using publicly available RNA-seq data. The RNA-seq data were downloaded from the NCBI Sequence Read Archive (SRA) database (4) by fasterq-dump (version 2.10.0) (https://github.com/ncbi/sra-tools) and preprocessed by fastp (version 0.20.1) (5) with options “-w 4,” “-y -3,” and “-x.” Generation of genome indexes and mapping analysis were conducted by STAR (version 2.7.8a) (6) with the default parameters. The expression levels were calculated as transcripts per million (TPM) using the formula: $TPM = \frac{RPKM}{\sum(RPKM)} * 10^6$. Reads per kilobase of exon per million mapped reads (RPKM) were calculated by featureCounts (version 2.0.1) (7). Mapped patterns of RNA-seq reads were visualized by IGV (version 2.9.4) (8). SRA Run accession, genome file, and gene annotation file for mapping analysis are shown in Dataset S1. Because RNA-seq data were unavailable, we did not investigate expression levels in the following tissues: brain (temporal lobe), colon, and lymph node in rhesus macaque; and testes and ovaries in crab-eating macaque and chimpanzee.

Expression of CTR1 and truncated *env* in crab-eating macaque. Total RNAs from testis and ovary in crab-eating macaque were obtained from Fukushima Medical University Medical-Industrial Translational Research Center (TR Center), and were used for cDNA synthesis. cDNA was synthesized using a PrimeScript II first-strand cDNA synthesis kit (Takara), in accordance with the manufacturer's instructions. CTR1 was amplified with SYBR Premix Ex Taq II (Tli RNaseH Plus; Takara) in a CFX96 Touch real-time PCR detection system (Bio-Rad) using primers hSLC31A1-rtF1 (5’-GCCTGCTGCGTAAGTCACAAG-3’) and hSLC31A1-rtR1 (5’-TGCAGGAGGTGAGGAAAGCTC-3’). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was amplified using primers hGAPDH-F-Tsu (5’-CACCACCATGGAGAAGGCTG-3’) and hGAPDH-R-Tsu (5’-GCTGATGATCTTGAGGCTGTTGT-3’). The expression of truncated *env* gene (CmRV-trunc1) was detected by PCR using KOD One PCR Master Mix Blue (Toyobo), using primers Cmtr-RT-F1(5'-CCACGCGGACATATGAAACCGCCCG-3') and Cmtr-RT-R1(5'-GGACCAACTATAGCAATACAGGTGAGT-3') for a 333bp-amplicon of CmRV-trunc1 or primers Cmtr-RT-F1 and Cmtr-RT-R2 (5'-CCAGTAGGATTGGTGGAATCAATGAAT-3') for a 920bp-amplicon of CmRV-trunc1. The thermal cycling conditions were: 35 cycles of amplification, consisting of denaturation at 96 °C for 10 s, annealing at 63 °C for 5 s, and extension at 68 °C for 1 s for the 333-bp amplicon, and denaturation at 96 °C for 10 s, annealing at 67 °C for 5 s, and extension at 68 °C for 1 s for the 920-bp amplicon. Agarose gel electrophoresis of the PCR products was conducted, and the gels were stained by Midori Green Xtra (Nippon Genetics, Bunkyo, Japan). The PCR products were confirmed by sequencing (Fasmac Co., Ltd., Atsugi, Japan). For quantitative RT-PCR of CmERV-trunc1in testis and ovary, cDNA was amplified with SYBR Premix Ex Taq II (Tli RNaseH Plus; Takara) in a CFX96 Touch real-time PCR detection system. CmRVtrunc1 was amplified using primers Cmtr-RT-F1 and Cmtr-RT-R1. Crab-eating macaque GAPDH was amplified using primers hGAPDH-F-Tsu and hGAPDH-R-Tsu.

Pseudotyped virus preparation. GPLac cells, *env*-negative packaging cells containing a LacZ-coding pMXs retroviral vector, were seeded in six-well plates 1 day prior to transfection. Cells were transfected with *env* expression plasmids to produce LacZ-carrying Env-pseudotyped virus. After 48 h, cell supernatants were collected, filtered through a 0.22-μm filter (Merck, Darmstadt Germany), and stored at −80°C. The following Env expression plasmids were used for pseudotyped virus preparations: pFUΔss clone33 (FeLV-A clone 33 env), pFUΔss GB (FeLV-B Gardner-Arnstein env) (kindly provided by Dr. Hajime Tsujimoto), pFUΔss TG35-2 (FeLV TG35-2 env), pFUΔss Ty2.0 (FeLV-D TY26 env), pFUΔss ON-T(FeLV-D ON-T env), pFUΔss ON-C (FeLV-D ON-C env), pFUΔss 44B (FeLV-D 44B env), pFUΔss DC8 (ERV-DC8 env), pFUΔss DC14 (ERV-DC14 env), pFUΔss DC19 (ERV-DC19 env), pFUΔss DC6 (ERV-DC6 env), pFUΔss DC10 (ERV-DC10 env), and pFUΔss 4070A (amphotropic MLV 4070A env) (1, 3, 9). Reconstituted functional ancestral Envs of ERV-DC7 and ERV-DC16 (i.e., ERV-DC7rec and ERV-DC16rec, respectively) containing the minimal required back mutations (R407G-I427N-T429A and D431Y, respectively) and missing the stop codon were previously described (10). ERV-DC7rec and ERV-DC16rec Env-pseudotyped viruses were prepared using a LacZ-coding pMSCV retroviral vector, rather than a pMXs retroviral vector. Previously described primate *env* genes (chimpanzee CERV1, CERV2, and rhesus macaque RhERV2-A) (11) were commercially synthesized (GENEWIZ). Genes containing C-terminal c-Myc tags were inserted into the pFUΔss vector. A DNA construct encoding a newly identified Env (CmRV1, CmRV1_1, CmRV1_2, and RmRV1) in this study from crab-eating macaques and rhesus macaques was commercially synthesized with a c-Myc tag at the 3' end (GENEWIZ). The gene was inserted into the pFUΔss vector.

Infection assay. Cells were inoculated with virus in 24-well plates containing 10 μg/mL polybrene (Santa Cruz Biotechnology, Santa Cruz, CA, USA) for 2 h. After the addition of fresh medium, cells were cultured for 2 days post-infection. To conduct the LacZ assay, culture supernatants were discarded and the cells were fixed with 250 μL of 2% glutaraldehyde for 15 min at room temperature, and then stained with 250 μL of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) solution. After incubation at 37°C for 2 h, nuclei in LacZ-positive cells were counted using a microscope and images were saved using a camera.

Viral interference assay. 293T/FeLV-D, 293T/FeLV-B, 293T/ERV-DC14, and 293T/ERV-DC10 cells were used as target cells in interference assays. Target cells were infected with each pseudotyped virus in the presence of polybrene for 2 days, and then stained with X-Gal, in accordance with the method described in the "*Infection assay*" subsection. Single-cycle infectivity was assessed by counting blue-stained nuclei.

Envelope gene synthesis. The following *env* genes were commercially synthesized (GENEWIZ): CERV1, CERV2, and RhERV2-A (11), and the following *env* genes with a c-Myc tag at the 3' end were CmRV1/RmRV1, CmRV1_1, CmRV1_2, CmRV-trunc1, PNRC2-ChRV, PNRC2-BoRV, and CERV2 provirus (ChrY); (Eurofins Genomics, Ohta Japan): BoRV-7 and GoRV-trunc1 (SI Figs. S5 and S15). These *env* genes were cloned into the pFUΔss expression plasmid.

Viral inhibition assay in the presence of truncated ERV Env proteins. The truncated Env proteins were prepared from the supernatants of HEK293T cells transfected with pFUΔss ERV-DC7 env, pFUΔss ERV-DC16 env, pFUΔss CmRV-trunc1 env, pFUΔss PNRC2-ChRV env, pFUΔss PNRC2-BoRV env, pFUΔss BoRV-7 env, pFUΔss GoRV-trunc1 env, pFUΔss CERV2 provirus (ChrY) env, or pFUΔss CERV2-trunc env. The supernatants were filtered through a 0.22-μm filter and stored at –80°C. The pFUΔss RmRV-trunc1 env was constructed by site-directed mutagenesis based on pFUΔss CmRV-trunc1 env using the following primers; Cmtr-mt-F1 (5'-

CCGTGGCTGTGGGCCCCAACCAAGTGCTGAGCG-3’) and Cmtr-mt-R1(5’- CGCTCAGCACTTGGTTGGGGCCCACAGCCACGG-3’), and Cmtr-mt-F2 (5’-CCTCCCCTCCTCTGATTAGCACCTCCAGCGAGTACAC-3’) and Cmtr-mt-R2 (5’-GTGTACTCGCTGGAGGTGCTAATCAGAGGAGGGGAGG-3’). CERV2-trunc (corresponding to amino acid positions 1 to 188 of CERV2) was artificially constructed from CERV2 by PCR using the PFU2S primer (5’-ACAGATCCAAGCTGTGACCG-3’) and CERV2env-tr2-R (5’-GTAGAATTCTCACAGATCCTCCTCGGAGATCAGCTTCTGCTCGATTAAACCGGTGGTCCATCTGG-3’). The infection assay was conducted as described above. Briefly, target cells were treated with 250 µL of cell supernatants containing 10 µg/mL polybrene, in 24-well plates; pseudotyped viruses were then incubated with target cells in the presence of 10 µg/mL polybrene. After the addition of fresh medium, the cells were stained with X-Gal at 2 days post-infection. Single-cycle infectivity was assessed by counting blue-stained nuclei under a microscope and images were saved using a camera.

RNA interference. MISSION® esiRNA (Sigma-Aldrich, Tokyo, Japan) was used to silence human CTR1 expression. Lipofectamine RNAiMAX transfection reagent (Invitrogen) was used to transfect cells with 400 ng of siRNAs targeting human CTR1 or EGFP as a control. The transfection was repeated 24 h later. The next day, cells were trypsinized and seeded in 24-well plates. Cells were infected with pseudotyped viruses the following day and an infection assay was conducted. The expression of CTR1 was evaluated by the following method. Total RNA was extracted from cells using an RNeasy Mini Kit (Qiagen), in accordance with the manufacturer's instructions. Thereafter, cDNA was synthesized using a PrimeScript II first-strand cDNA synthesis kit (Takara), in accordance with the manufacturer's instructions. Prior to reverse transcription, RNA samples were treated with RNase-Free DNase Set (Qiagen). cDNA was amplified with SYBR Premix Ex Taq II (Tli RNaseH Plus; Takara) in a CFX96 Touch real-time PCR detection system (Bio-Rad). Human CTR1 was amplified using primers hSLC31A1-rtF1 and hSLC31A1-rtR1. Human glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was amplified using primers hGAPDH-F-Tsu and hGAPDH-R-Tsu.

Receptor screening. The retroviral HEK293 cDNA library constructed in the pMX retroviral vector was prepared as described previously (12). Poly(A)$^+$ RNA of HEK293 cells was prepared using FastTrack kit ver. 2.0 (Thermo Fisher Scientific), in accordance with the manufacturer’s protocol. cDNA was synthesized from the RNA with oligo dT$_{12\text{-}18}$ primers, random hexamers, and Superscript II Reverse Transcriptase (Invitrogen). Double-stranded cDNA was then synthesized using DNA ligase, DNA polymerase I, and RNase H. Blunt-end cDNA was adapted with *Bst*XI adaptor (Invitrogen) and ligated to the *Bst*XI sites of pMX vector. The ligated DNA was amplified in DH10B competent cells and plasmid DNA was extracted using the Qiagen plasmid midi kit (QIAGEN). MDTF cells were transduced with the retroviral HEK293 cDNA library using the Retrovirus Packaging Kit Ampho (Takara). PLAT-GP cells stably introduced with MLV Gag-Pol were transfected with FeLV-D Ty2.0 Env (FeLV-D TY26) plasmid and either pMSCV-puro (Clontech, Mountain View, CA, USA) or pMXs encoding GFP using TransIT®-293 reagent. The supernatants were collected 2 days later, then filtered (0.22 µm) to prepare viral solutions. FeLV-D Env-pseudotyped virus was concentrated using Amicon Ultra-15 10K Centrifugal Filter Devices (Merck). The viruses were stored at −80°C until use. The viral titers were $>5\times10^5$ IU/mL. The FeLV-D (TY26) Env-pseudotyped virus was used to challenge MDTF cells transduced with the cDNA library in eight 10-cm culture dishes; cells were selected with 6 µg/mL puromycin (InvivoGen). Cells were then infected with FeLV-D Env-pseudotyped virus containing a GFP-coding pMXs retroviral vector. The chromosomal DNAs from GFP-positive cells were extracted with a DNeasy kit (Qiagen) and used as a PCR template with primers located in the pMXs vector cloning sites (MX-1S: 5’-GCATCGCAGCTTGGATACAC-3’ and MX-1R: 5’-TATTTTATCGTCGACCACTGTGC-3’). The PCR products were directly sequenced (Fasmac Co., Ltd.).

Detection of CTR1 expression by flow cytometry. Cells were harvested with Enzyme Free Cell Dissociation Solution PBS Based (Sigma-Aldrich) and washed with PBS containing 0.5% bovine serum albumin (0.5% PBSA). Mouse anti-HA antibody conjugated with fluorescein isothiocyanate

(Miltenyi Biotec, Bergisch Gladbach, Germany) was used to detect CTR1. Cells were incubated with 0.5% PBSA containing antibody on ice for 15 min, then washed with 0.5% PBSA. Finally, cells were analyzed with a BD Accuri C6 flow cytometer (BD Biosciences, Franklin Lakes, NJ, USA).

Detection of huCTR2 expression by RT-PCR. Total RNA was extracted from cell lines using the RNeasy Mini Kit (Qiagen) and RNase-Free DNase Set (Qiagen). cDNA was synthesized using the PrimeScript II first-strand cDNA synthesis kit (Takara). PCR for detecting huCTR2 in cell lines was performed using KOD One PCR Master Mix Blue (Toyobo) and primers Myc-HuCTR2 F2 (5'-AAGCTGATCAGCGAGGAGG-3') and Myc-HuCTR2 R2 (5'-AACTTGGTTCAGCAGCTTGG-3'). Thermal cycling conditions were as follows: 30 cycles of 98°C for 10 s, 59.5°C for 5 s, and 68°C for 1 s. The PCR products were electrophoresed on a 1% agarose gel and stained with ethidium bromide.

Immunoprecipitation and immunoblotting. HEK293T cells were transfected with plasmids using TransIT®-293 reagent in six-well plates. The cell supernatants were collected 2 days after transfection. Cell supernatants were filtered through a 0.22-µm filter and immunoprecipitated with anti-c-Myc monoclonal antibody agarose beads (Takara) at 4°C overnight. Sample Buffer Solution with Reducing Reagent (6×) (Nacalai Tesque, Kyoto, Japan) was added to the purified protein and heated at 100°C for 5 min. SDS-PAGE was performed with 10%–20% gels (Invitrogen) at 100 V for 2 h. Western blotting was performed with anti-c-Myc monoclonal antibody conjugated with horseradish peroxidase (FUJIFILM Wako Pure Chemical Corporation). LumiGLO® Reagent (20×) and 20× peroxide (Cell Signaling Technology, Danvers, MA, USA) were used as the substrate; blots were imaged using the Lumino Image Analyzer LAS2000 (Fujifilm, Minato city, Japan) or Amersham ImageQuant 800 (Cytiva, Shinjuku, Japan).

Phylogenetic and sequence analyses.

To identify primate ERV-related viruses, the amino acid sequence of CmRV1 was first used in a protein BLAST search of the NR- database (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Next, nucleotide sequences were used to obtain their features via a BLAT search of UCSC primate genomes including chimpanzee (Clint_PTRv2/panTro6), bonobo (Mhudiblu_PPA_v0/panPan3), crab-eating macaque (Macaca_fascularis_5.0/macFas5), rhesus macaque (Mul_10/rheMac10), and gorilla (Kamilah_GGO_v0/gorGor6). The multiple sequence alignment (MSA) using gammaretroviral Env proteins was constructed by MAFFT (version 7.475) (13) with the option "--auto" (SI Fig. S7).

The MSA using ERV-DC nucleotide sequences (SI Fig. S1) was constructed by MAFFT with the option "--auto." MSAs using amino acid sequences of *gag, pol, and env* genes were constructed by MAFFT with the options "--add" and "--keeplength." Phylogenetic trees of *gag* and *pol* genes were constructed by the maximum likelihood method using IQTREE (version 1.6.12) (14). Substitution models were selected based on the BIC score provided by ModelFinder (15): TPM2+R3 for ERV-DCs (SI Fig. S1), JTTDCMut+F+G4 for *gag* genes (SI Fig. S10A), and LG+F+R4 for *pol* genes (SI Fig. S10B). The branch supportive values were measured using an ultrafast bootstrap method by UFBoot2 (16) with 1,000 replicates. The *env* phylogenetic tree was constructed based on the neighbor-joining method using the JTT model (17). Tree robustness was evaluated using the bootstrap method (1,000 replicates). The programs used to create the *env* tree (SI Fig. S10C) were implemented in MEGA X (18). Sequence accession numbers used for phylogenetic analyses are shown in Dataset S2. Tree visualization was performed by the ggtree package (version 2.2.1) (19).

Estimation of the integration timings of primate truncated Envs and full-length Envs. The integration timings of primate ERVs were estimated by the ortholog-dating method (20). First, *env* genes of CERV-related viruses were searched in the NCBI RefSeq genomic database (version: 20210511) by tBLASTn (version 2.5.0). Next, 10 kbp upstream and downstream genomic regions of CERV-like viral *env* genes were extracted by seqkit (version 0.15.0) (21); sequence similarities were compared by BLASTN (version 2.5.0) with the option "-evalue 1E-10." The presence/absence patterns of orthologous *env* loci were determined based on alignment coverage between flanked sequences of *env* genes. The alignment coverage among orthologous ERVs was 99.87% (PNCR2-ChRV and PNCR2-BoRV), 92.38% (CmRV-trunc1 and RmRV-trunc1), and 86.61% (CmRV1_2 and its ortholog in rhesus macaque) (SI Fig. S11). The sequence alignment of these orthologs was constructed by MAFFT (SI Figs. S12–S14). The species divergence time during primate evolution was determined using the Timetree database (22).

Ethical approval. Animal studies were conducted in accordance with the Guidelines for the Care and Use of Laboratory Animals of the Ministry of Education, Culture, Sports, Science and Technology, Japan. All experiments were approved by the Genetic Modification Safety Committee of Yamaguchi University, Yamaguchi, Japan.

Data availability. The relevant codes and data are available at https://github.com/Junna-Kawasaki/Refrex_2021. All other study data are included in the article and/or supporting information.

SI References

1. Y. Anai *et al.*, Infectious Endogenous retroviruses in cats and emergence of recombinant viruses. *J. Virol.* **86**, 8634-8644 (2012).
2. K. Kuse *et al.*, Existence of two distinct infectious endogenous retroviruses in domestic cats and their different strategies for adaptation to transcriptional regulation. *J. Virol.* **90**, 9029-9045 (2016).
3. J. Ito *et al.*, Refrex-1, a soluble restriction factor against feline endogenous and exogenous retroviruses. *J. Virol.* **87**, 12029-12040 (2013).
4. R. Leinonen, H. Sugawara, M. Shumway, o. b. o. t. I. N. S. D. Collaboration, the sequence read archive. *Nucleic Acids Res.* **39**, D19-D21 (2010).
5. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* **34**, i884-i890 (2018).
6. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* **29**, 15-21 (2012).
7. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* **30**, 923-930 (2013).
8. J. T. Robinson, H. Thorvaldsdóttir, A. M. Wenger, A. Zehir, J. P. Mesirov, Variant Review with the Integrative Genomics Viewer. *Cancer Res.* **77**, e31-e34 (2017).
9. A. Miyake *et al.*, Reduced folate carrier: an entry receptor for a novel feline leukemia virus variant. *J. Virol.* **93**, e00269-00219 (2019).
10. J. Ito, T. Baba, J. Kawasaki, K. Nishigaki, Ancestral mutations acquired in refrex-1, a restriction factor against feline retroviruses, during its cooption and domestication. *J. Virol.* **90**, 1470-1485 (2016).
11. S. J. Soll, S. J. D. Neil, P. D. Bieniasz, Identification of a receptor for an extinct virus. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19496-19501 (2010).
12. T. Kitamura, Y. Morikawa, "Isolation of T-Cell Antigens by Retrovirus-Mediated Expression Cloning" in T Cell Protocols. Methods in Molecular Biology™*,* K. P. Kearse, Ed. (Humana Press, Totowa, NJ, 2000), vol. 134, pp. 143-152.
13. K. Katoh, D. M. Standley, MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).
14. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268-274 (2014).
15. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods.* **14**, 587-589 (2017).
16. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518-522 (2017).
17. D. T. Jones, W. R. Taylor, J. M. Thornton, The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275-282 (1992).
18 S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547-1549 (2018).
19. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28-36 (2017).
20. P. Aiewsakun, A. Katzourakis, Endogenous viruses: Connecting recent and ancient viral evolution. *Virology.* **479-480**, 26-37 (2015).
21. W. Shen, S. Le, Y. Li, F. Hu, SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One.* **11**, e0163962 (2016).
22. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812-1819 (2017).

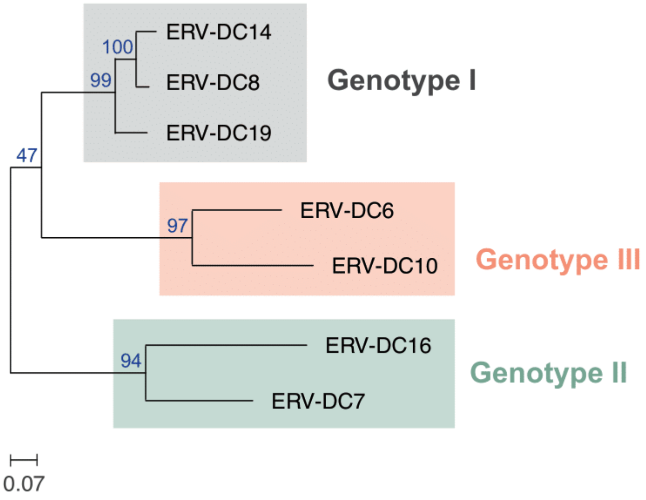
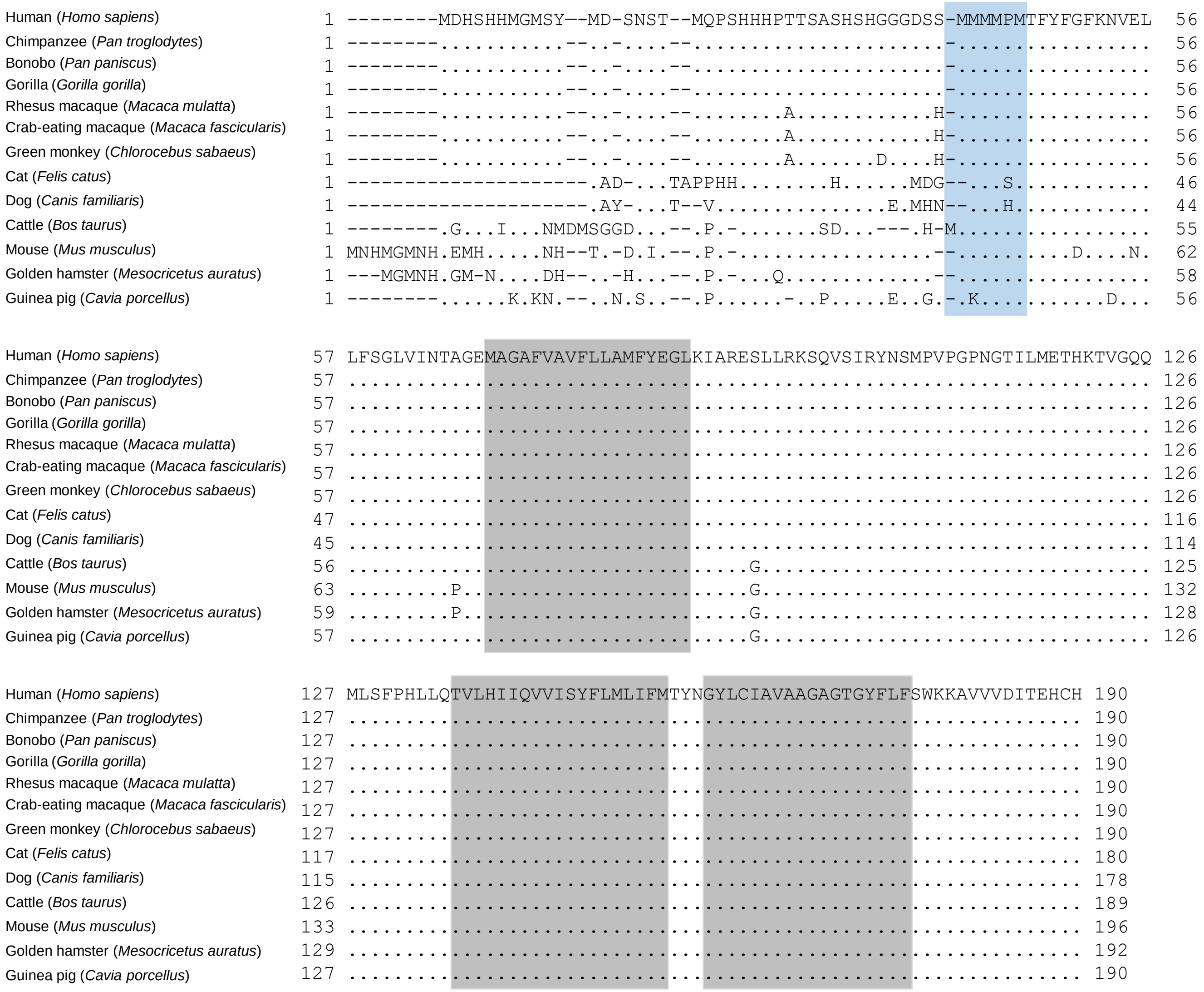


Figure S1. ERV-DC genotypes. The phylogenetic tree was constructed using nucleotide sequences of ERV-DCs based on the maximum likelihood method. Scale bars indicate the genetic distance (substitutions per site). The blue labels on branches indicate the bootstrap support values (%) with 1,000 replicates. The colored boxes highlight ERV-DC genotypes I (gray), II (green), and III (pink).



```
Human (Homo sapiens)                           1 --------MDHSHHMGMSY--MD-SNST--MQPSHHHPTTSASHSHGGGDSS-MMMMPMTFYFGFKNVEL  56
Chimpanzee (Pan troglodytes)                   1 --------...........--..-....--......................-.................  56
Bonobo (Pan paniscus)                          1 --------...........--..-....--......................-.................  56
Gorilla (Gorilla gorilla)                      1 --------...........--..-....--......................-.................  56
Rhesus macaque (Macaca mulatta)                1 --------...........--..-....--........A............H-.................  56
Crab-eating macaque (Macaca fascicularis)      1 --------...........--..-....--........A............H-.................  56
Green monkey (Chlorocebus sabaeus)             1 --------...........--..-....--........A.......D....H-.................  56
Cat (Felis catus)                              1 ---------------------.AD-...TAPPHH........H......MDG--...S............  46
Dog (Canis familiaris)                         1 ---------------------.AY-...T--V...............E.MHN--...H............  44
Cattle (Bos taurus)                            1 --------.G...I...NMDMSGGD...--.P.-.......SD...---.H-M.................  55
Mouse (Mus musculus)                           1 MNHMGMNH.EMH.....NH--T.-D.I.--.P.-.................--..........D....N.  62
Golden hamster (Mesocricetus auratus)          1 ---MGMNH.GM-N....DH--..-H...--.P.-...Q.............--.................  58
Guinea pig (Cavia porcellus)                   1 --------......K.KN.--..N.S..--.P......-..P.....E..G.-.K...........D...  56

Human (Homo sapiens)                          57 LFSGLVINTAGEMAGAFVAVFLLAMFYEGLKIARESLLRKSQVSIRYNSMPVPGPNGTILMETHKTVGQQ 126
Chimpanzee (Pan troglodytes)                  57 ...................................................................... 126
Bonobo (Pan paniscus)                         57 ...................................................................... 126
Gorilla (Gorilla gorilla)                     57 ...................................................................... 126
Rhesus macaque (Macaca mulatta)               57 ...................................................................... 126
Crab-eating macaque (Macaca fascicularis)     57 ...................................................................... 126
Green monkey (Chlorocebus sabaeus)            57 ...................................................................... 126
Cat (Felis catus)                             47 ...................................................................... 116
Dog (Canis familiaris)                        45 ...................................................................... 114
Cattle (Bos taurus)                           56 ...................................G.................................. 125
Mouse (Mus musculus)                          63 .........P.........................G.................................. 132
Golden hamster (Mesocricetus auratus)         59 .........P.........................G.................................. 128
Guinea pig (Cavia porcellus)                  57 ...................................G.................................. 126

Human (Homo sapiens)                         127 MLSFPHLLQTVLHIIQVVISYFLMLIFMTYNGYLCIAVAAGAGTGYFLFSWKKAVVVDITEHCH 190
Chimpanzee (Pan troglodytes)                 127 ................................................................ 190
Bonobo (Pan paniscus)                        127 ................................................................ 190
Gorilla (Gorilla gorilla)                    127 ................................................................ 190
Rhesus macaque (Macaca mulatta)              127 ................................................................ 190
Crab-eating macaque (Macaca fascicularis)    127 ................................................................ 190
Green monkey (Chlorocebus sabaeus)           127 ................................................................ 190
Cat (Felis catus)                            117 ................................................................ 180
Dog (Canis familiaris)                       115 ................................................................ 178
Cattle (Bos taurus)                          126 ................................................................ 189
Mouse (Mus musculus)                         133 ................................................................ 196
Golden hamster (Mesocricetus auratus)        129 ................................................................ 192
Guinea pig (Cavia porcellus)                 127 ................................................................ 190
```

Figure S2. Alignment of the amino acid sequences of CTR1. Dots indicate conserved amino acid residues, while hyphens indicate gaps in the amino acid sequence. Transmembrane domains are shown in gray. The MMMMXM copper-coordination motif is shown in blue. The feline CTR1 sequence was determined by cloning the gene from feline FT-1 cell lines. All other sequences were obtained from the National Center for Biotechnology Information, with the following gene accession numbers: human (NM_001859.4), chimpanzee (XM_520197.6), bonobo (XM_003833052.3), gorilla (XM_019033403.2), rhesus macaque (NM_001257506.1), crab-eating macaque (XM_005581024.2), green monkey (XM_007968369.2), cat (XM_023242310.1), dog (XM_038682637.1), cattle (NM_001100381.1), mouse (NM_175090.4), golden hamster (XM_040757395.1), and guinea pig (XM_003463793.4).

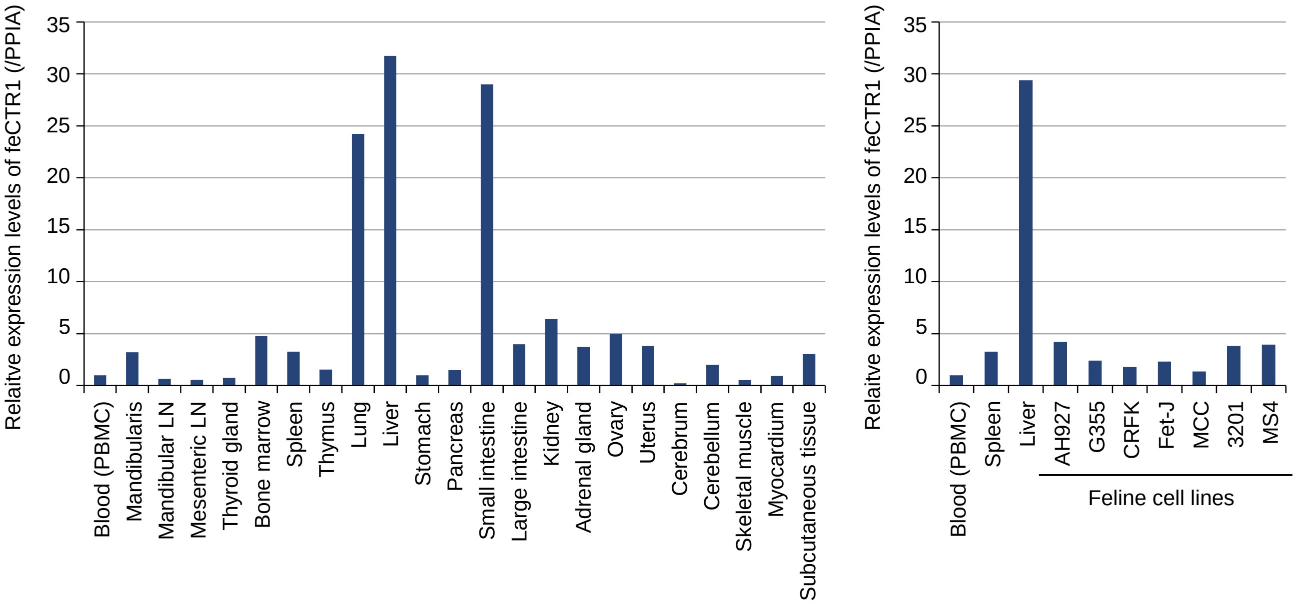
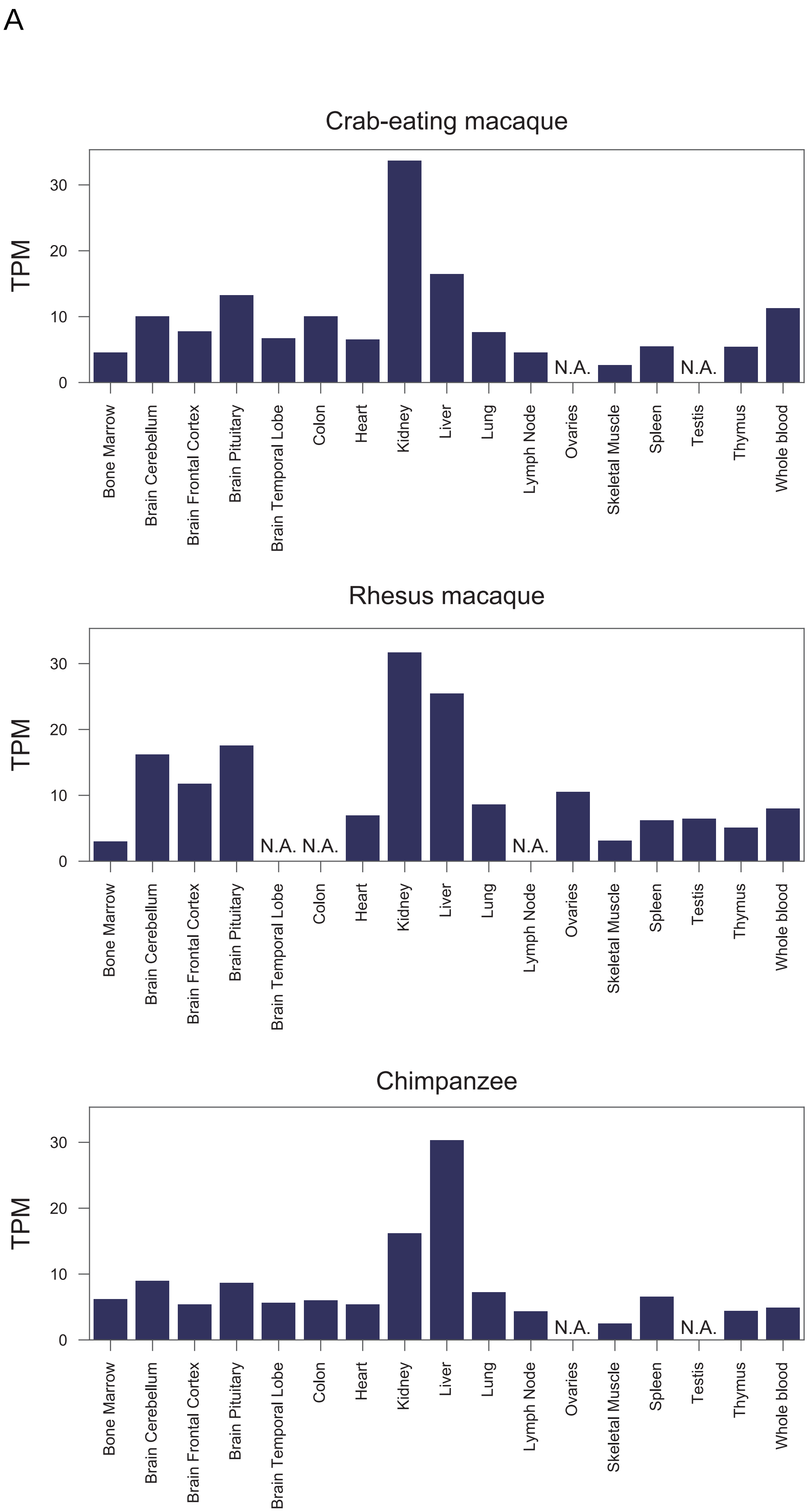
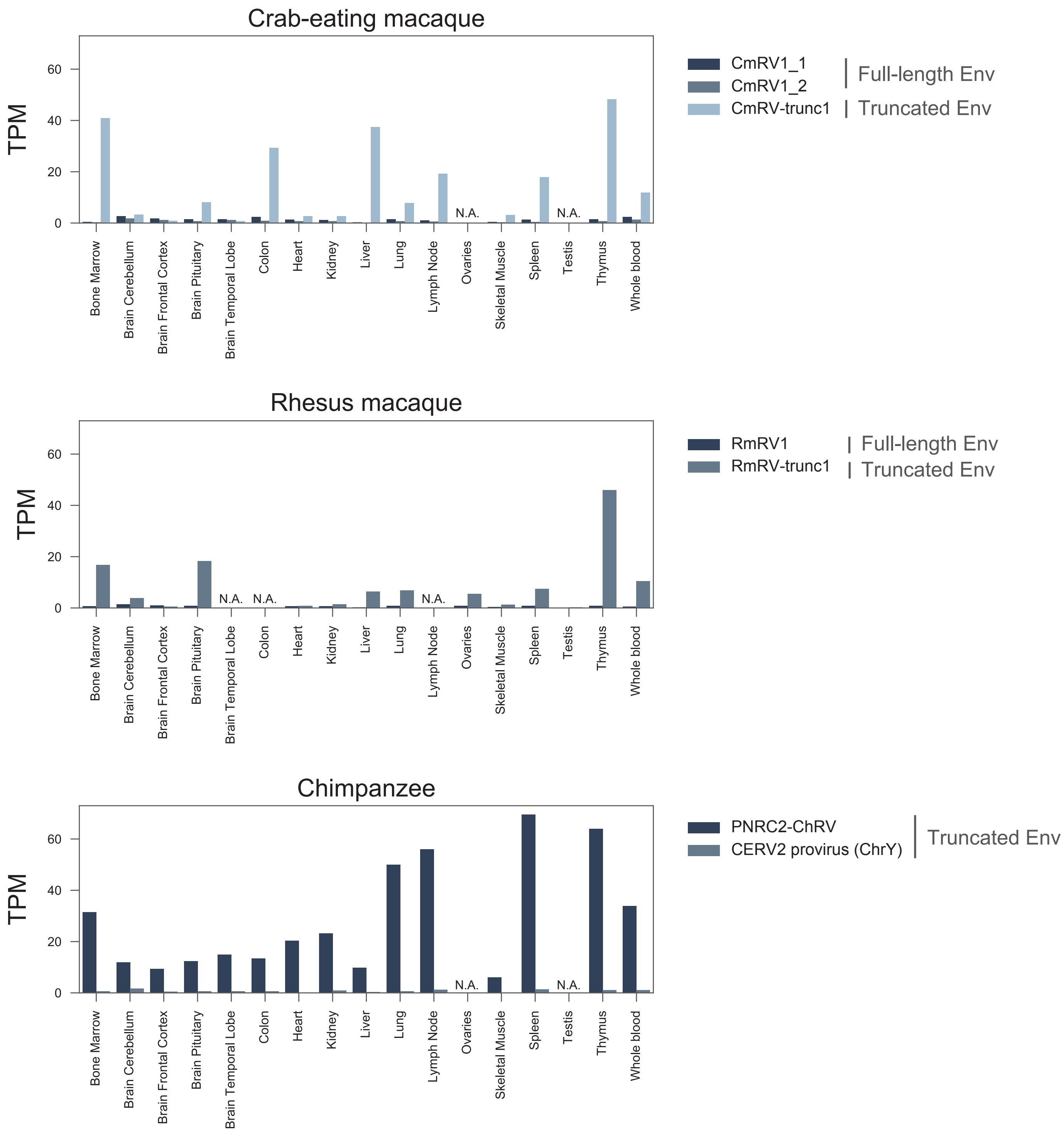


Figure S3. CTR1 expression in feline tissues and cell lines. Quantification of feline CTR1 transcripts by quantitative RT-PCR in feline tissues and feline cell lines. The y-axis indicates expression level normalized to the expression of peptidylprolyl isomerase A (PPIA). LN, Lymph node.


A
Crab-eating macaque
TPM
0
10
20
30
N.A.
N.A.
Bone Marrow
Brain Cerebellum
Brain Frontal Cortex
Brain Pituitary
Brain Temporal Lobe
Colon
Heart
Kidney
Liver
Lung
Lymph Node
Ovaries
Skeletal Muscle
Spleen
Testis
Thymus
Whole blood
Rhesus macaque
TPM
0
10
20
30
N.A.
N.A.
N.A.
Bone Marrow
Brain Cerebellum
Brain Frontal Cortex
Brain Pituitary
Brain Temporal Lobe
Colon
Heart
Kidney
Liver
Lung
Lymph Node
Ovaries
Skeletal Muscle
Spleen
Testis
Thymus
Whole blood
Chimpanzee
TPM
0
10
20
30
N.A.
N.A.
Bone Marrow
Brain Cerebellum
Brain Frontal Cortex
Brain Pituitary
Brain Temporal Lobe
Colon
Heart
Kidney
Liver
Lung
Lymph Node
Ovaries
Skeletal Muscle
Spleen
Testis
Thymus
Whole blood

B


Crab-eating macaque
TPM
0
20
40
60
N.A.
N.A.
CmRV1_1
CmRV1_2
CmRV-trunc1
Full-length Env
Truncated Env
Rhesus macaque
TPM
0
20
40
60
N.A.
N.A.
N.A.
RmRV1
RmRV-trunc1
Full-length Env
Truncated Env
Chimpanzee
TPM
0
20
40
60
N.A.
N.A.
PNRC2-ChRV
CERV2 provirus (ChrY)
Truncated Env
Bone Marrow
Brain Cerebellum
Brain Frontal Cortex
Brain Pituitary
Brain Temporal Lobe
Colon
Heart
Kidney
Liver
Lung
Lymph Node
Ovaries
Skeletal Muscle
Spleen
Testis
Thymus
Whole blood

C

Sample		Quantification cycle (Cq) values		
		CTR1	CmRV-trunc1	GAPDH
Crab-eating macaque	Ovary	22.80	25.05	17.17
	Testis	24.24	26.62	17.57
293T cells		19.99	0	13.43

Figure S4. Expression of CTR1, and truncated and full-length *env* genes in primate tissues. (A) CTR1 expression levels and **(B)** expression levels of truncated and full-length *env* genes were analyzed using publicly available RNA-seq data from crab-eating macaque (upper), rhesus macaque (middle), and chimpanzee (lower). The x-axis indicates tissues in which expression levels were investigated, and the y-axis indicates the normalized expression levels of CTR1 (A) and *env* genes (B), as transcripts per million (TPM). N.A.: expression level in the indicated tissue was not investigated because RNA-seq data were not available. **(C)** Expression of CTR1 and truncated *env* (CmRV-trunc1) was analyzed using cDNAs from ovary and testis from crab-eating macaque by quantitative RT-PCR (qRT-PCR). Quantification of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is also shown. 293T cells were used as a control. The data represent the quantification cycle (Cq) values.

```
                              Signal peptide                         ┌→ SU
RmRV1 (chr11)            1 METPTLRRPAQDKTLSYLAPCLLLALLPCVAGSNNPHRPYNLTWQVIDFSTHEVLDKTSK  60
RmRV1 (KT330634.1)       1 ............................................................  60
CmRV1 (KT330768.1)       1 ............................................................  60
CmRV1 (KT330663.1)       1 ............................................................  60
CmRV1_1 (chr13)          1 .......G....................................................  60
CmRV1_2 (chr14)          1 .......G.......................D............................  60

                                                   VRA
                                      ┌──────────────────────────────────────────────┐
RmRV1 (chr11)           61 IAPMGTWFPDLYFNLDKIAKIDDMEGGEWRKQARRVSVSRNGFYACPGFRTGEMKKTCGE 120
RmRV1 (KT330634.1)      61 ............................................................ 120
CmRV1 (KT330768.1)      61 ............................................................ 120
CmRV1 (KT330663.1)      61 ............................................................ 120
CmRV1_1 (chr13)         61 ......................................................R..... 120
CmRV1_2 (chr14)         61 ......................................................R..... 120

                                                              VRB
                                                         ┌──────────┐
RmRV1 (chr11)          121 IDALFCASWSCITTNDGEWKWATKPWYITMSFVQRCTRTRYSKTCNLVRIKFEDAAKSDN 180
RmRV1 (KT330634.1)     121 ............................................................ 180
CmRV1 (KT330768.1)     121 ............................................................ 180
CmRV1 (KT330663.1)     121 ............................................................ 180
CmRV1_1 (chr13)        121 ............................................................ 180
CmRV1_2 (chr14)        121 ..................................C......................... 180

                                                               ┌→ PRR
RmRV1 (chr11)          181 RWISGLIWGLYLYQKPLYGIPIQIKLIVNPITAPVAVGPNQVLSETRKPLVPAPREPQPR 240
RmRV1 (KT330634.1)     181 ............................................................ 240
CmRV1 (KT330768.1)     181 ............................................................ 240
CmRV1 (KT330663.1)     181 ............................................................ 240
CmRV1_1 (chr13)        181 ............................................................ 240
CmRV1_2 (chr14)        181 ............................................K............... 240

                                                  PRR ←┐
RmRV1 (chr11)          241 APKSTSPPLISTSSKYTPSAQTVTRGPLNLGIGDRLLNLIKGSYFALNQTKPEFTSSCWL 300
RmRV1 (KT330634.1)     241 ............................................................ 300
CmRV1 (KT330768.1)     241 ............................................................ 300
CmRV1 (KT330663.1)     241 ............................................................ 300
CmRV1_1 (chr13)        241 ............................................................ 300
CmRV1_2 (chr14)        241 .....................N......D............................... 300

RmRV1 (chr11)          301 CLATGPPYYEGIASTNNFTNSANPTGCAWEQQRKLTLAEVSGSGTCIGQVPPSHQHLCNV 360
RmRV1 (KT330634.1)     301 ............................................................ 360
CmRV1 (KT330768.1)     301 ............................................................ 360
CmRV1 (KT330663.1)     301 ............................................................ 360
CmRV1_1 (chr13)        301 ............................................................ 360
CmRV1_2 (chr14)        301 ............................................................ 360

RmRV1 (chr11)          361 TLTVPSSNHYLVPSETDWWACNTGLTPCISTAVFSSGTHYCVLVQVVPRVYYHSGDSFDL 420
RmRV1 (KT330634.1)     361 ............................................................ 420
CmRV1 (KT330768.1)     361 ............................................................ 420
CmRV1 (KT330663.1)     361 ............................................................ 420
CmRV1_1 (chr13)        361 ............................................................ 420
CmRV1_2 (chr14)        361 ............................V............................... 420

                              SU ←─┬─→ TM
RmRV1 (chr11)          421 RYEQKTHTRPKREPISLTLAVMLGIGVAAGVGTGTAALVHGNHHLQQLRVAIDEDLRAIE 480
RmRV1 (KT330634.1)     421 ............................................................ 480
CmRV1 (KT330768.1)     421 ............................................................ 480
CmRV1 (KT330663.1)     421 ............................................................ 480
CmRV1_1 (chr13)        421 ............................................................ 480
CmRV1_2 (chr14)        421 ............................................................ 480

RmRV1 (chr11)          481 QSITKLEESLTSLSEVVLQNRRGLEIVFLKEGGLCAALKEQCCFYADHSGVVKDSMAKLR 540
RmRV1 (KT330634.1)     481 ............................................................ 540
CmRV1 (KT330768.1)     481 ............................................................ 540
CmRV1 (KT330663.1)     481 ............................................................ 540
CmRV1_1 (chr13)        481 ............................................................ 540
CmRV1_2 (chr14)        481 ............................................................ 540

RmRV1 (chr11)          541 ERLDKRKKERESQQNWFENWYNQSPWLSTLISTILGPLILLMLILTFGPCILNRLLTLIK 600
RmRV1 (KT330634.1)     541 ............................................................ 600
CmRV1 (KT330768.1)     541 ............................................................ 600
CmRV1 (KT330663.1)     541 ............................................................ 600
CmRV1_1 (chr13)        541 .........................................T.................. 600
CmRV1_2 (chr14)        541 .........................................T.................. 600

                           TM ←─┬─→ R peptide
RmRV1 (chr11)          601 NRLNIVHAMVLTQQYQTLRTEEEAQD 626
RmRV1 (KT330634.1)     601 .......................... 626
CmRV1 (KT330768.1)     601 .......................... 626
CmRV1 (KT330663.1)     601 .......................... 626
CmRV1_1 (chr13)        601 .......................... 626
CmRV1_2 (chr14)        601 .......................... 626
```

Figure S5. Predicted RmRV1, CmRV1, CmRV1_1, and CmRV1_2 Env proteins. Primate ERV Env proteins from crab-eating macaque and rhesus macaque are shown using single-letter amino acid codes. The names of six ERV Envs are indicated with the gene accession number or chromosome number, as shown in SI Table S1. Dots, conserved amino acid residues. The CmRV1/RmRV1, CmRV1_1, and CmRV1_2 *env* genes encoding the indicated amino acids were DNA-synthesized. SU/TM and TM/R peptide boundaries and VRA/VRB/PRR were predicted based on homology with the RhERV2-A sequence (11). SU, surface unit; TM, transmembrane domain; VR, variable region; PRR, proline-rich region.

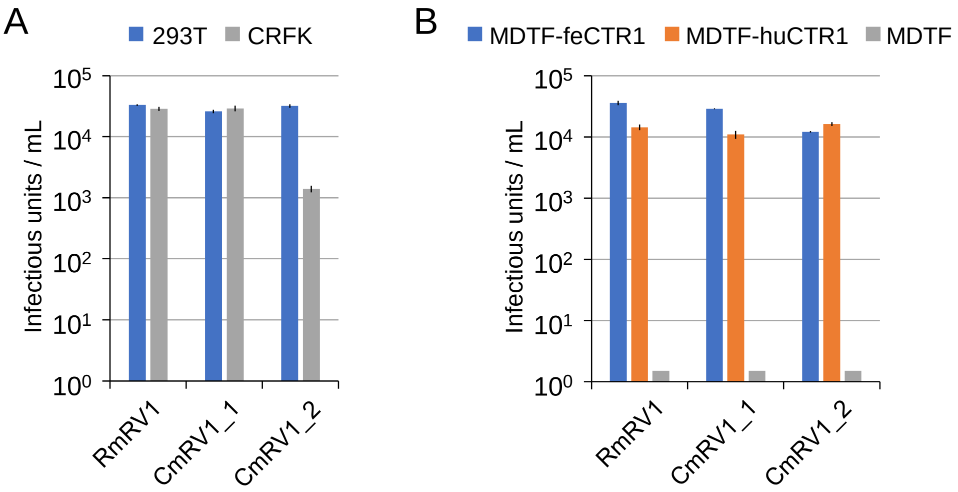


Figure S6. Infectivity of primate CmRV1_1, CmRV1_2, and RmRV1 (CmRV1). Infection of LacZ-coding primate Env-pseudotyped viruses in **(A)** HEK293T and CRFK cells and **(B)** MDTF-feCTR1, MDTF-huCTR1, and MDTF cells. The x-axis indicates the pseudotyped viruses and the y-axis indicates infectious units per milliliter of virus. Infective titers with standard deviations are means of three independent infection experiments.

```
                                     Signal peptide                                  SU
FeLV-A                 1   MESPTHPKPSKDKTLSWNLVFLV-GILFTIDIGMAN----------------PSPHQIYNVTWVITNVQTNTQANATSML 63
FeLV-B                 1   .......................-............----------------.....V.....T...LV.G.K....... 63
FeLV-D (TY26)          1   .------..PAGMVFL.V.TS.GA..GAK.---VKE----------------GN...V.TL..Q.YS-.SGEVVWEVQGN 54
FeLV-D (ON-T)          1   .------..PAGMVFL.V.TS.GA..GAK.---VKE----------------GN...V.TL..Q.YS-.SGEVVWEVQGN 54
FeLV-D (ON-C)          1     ------.PAGMVFL.V.TS.GA..GAK.---VKE----------------GN...V.TL..Q.YS-.SGEVVWEVQGN 52
FeLV-D (44B)           1   .------..PAGMVFL.V.TS.GA..GAK.---VKE----------------GN...V.TL..Q.YS-.SGEVVWEVQGN 54
ERV-DC10               1   .------R..ARMVFL.V.TS.GAEVGAR.---AKE----------------GN...V.TL..Q.YS-.SGEVVWEVQGN 54
ERV-DC14               1   .------..PAGMVFL.V.TS.GA..GAK.---VKE----------------GN...V.TL..Q.YS-.SGEVVWEVQGN 54
ERV-DC7 (Refrex-1)     1   .------..PTGMVFL.A.TS.GA..GAKN---VGE----------------GN...V.TL..QVYS-.SGEVVWEVQGN 54
ERV-DC16 (Refrex-1)    1   .------.LPTGMIFL.A.TS.GA..GAKN---VGK----------------GN...V.TL..QVYS-.SGEVVWEVQGN 54
RmRV1                  1   ..T..LRR.AQ.....YLAPC.L--LAL-LPCVAGS----------------NN..RP..L..QVIDFS.HEVLDK..KI 61
CmRV1                  1   ..T..LRR.AQ.....YLAPC.L--LAL-LPCVAGS----------------NN..RP..L..QVIDFS.HEVLDK..KI 61
CmRV1_1                1   ..T..LRG.AQ.....YLAPC.L--LAL-LPCVAGS----------------NN..RP..L..QVIDFS.HEVLDK..KI 61
CmRV1_2                1   ..T..LRG.AQ.....YLAPC.L--LAL-LPCVADS----------------NN..RP..L..QVIDFS.HEVLDK..KI 61
CERV2 (extinct)        1   ..TSSLQG.PQ....PCLTPH.L--LALFLPWVIGS----------------SN...P.RL..Q...FE.HEVL.E..HV 62
RhERV2-A (extinct)     1   ..T..LRG.AQ.....YLAPC.L--LAL-LPCVAGS----------------NN..RP..L..QVIDFS.HEVLDK..KI 61
CmRV-trunc1            1   ..T..LQG.AQ.....YLTPC.L--LAL-LPCAAGSNNPRGHMKPPGYIFIDNN..RP..L..QVI.FNNHEVLGE..KI 77
RmRV-trunc1            1   ..T..LQG.AQ.....YLTPC.L--LAL-LPCAAGSNNPRGHMKPPGYIFIDNN..RP..L..QVI.FNNHEVLGE..KI 77
PNRC2-ChRV             1   ..TSSLQG.PQ...FPCLTPH.L--LTLFLPWVVGS----------------SN...P.RL..Q...FK.HEVL.E..HV 62
PNRC2-BoRV             1   ..TSSLQG.PQ...FPCLTPH.L--LTLFLPWVVGS----------------SN...P.RL..Q...FK.HEVL.E..HV 62
BoRV-7                 1   ..TSSLQG.PQ...RPCLTPH.L--LALFLPWVIGS----------------SN...P.RL..Q...FE.HEVL.E..HV 62
CERV2 provirus (ChrY)  1   ..TSSLQG.PQ....PCLTPH.L--LALFLPWVIGS----------------SN...P.RL..Q...FE.HEVL.E..HI 62
CERV2-trunc(artificial)1   ..TSSLQG.PQ....PCLTPH.L--LALFLPWVIGS----------------SN...P.RL..Q...FE.HEVL.E..HV 62
GoRV-trunc1            1   ..TSSLRG.PQ....PCLTPH.LLTLALFLPWVTG.----------------GN.Y.P.QL..QV..FE.HEVL.E..HV 64
                                                                  SU
                                                                           VRA
FeLV-A                 64  GTLTDVYPTLHVDLCDLVG--DTWEPIVLSPTNVKH---GARYPSSKYGCKTTDRKKQQQTYPFYVCPGHAPSLGPKGTH 138
FeLV-B                 64  .....AF..MYF....II.--N..N.-------------SDQE.FPG...DQPM.RW..RNT.........-----NRKQ 123
FeLV-D (TY26)          55  HA.NTWW.P.TP.F.Q.AAGL...DIPAR..K.LQSYMGERIQQMTAH..SSPTARCRLAQAE.....RDNRD-RATAHR 133
FeLV-D (ON-T)          55  HA.NTWW.P.TP.F.Q.AAGL...DIPAR..K.LQSYMGERIQQMTAH..SSPTARCRLAQAE.....RDNRN-RATAHR 133
FeLV-D (ON-C)          53  HA.NTWW.P.TP.F.Q.AAGL...DIPAR..E.LQSYMGERIQQMTAH..SSPTARCRLAQAE.....RDNRN-RATAHR 131
FeLV-D (44B)           55  HA.NTWW.P.TP.F.Q.AAGL...DIP.R..K.LQSYMGERIQQMTAP..SSPTARCRLAQAE.....RDNRD-RATAHR 133
ERV-DC10               55  HA.NTWW.P.TP.F.Q.AAGL...DIPDR..K.LETSME.TSQQLTPQ..SKPWKRCALAERD.....RDNRN-RATAHR 133
ERV-DC14               55  HA.NTWW.P.TP.F.Q.AAGL...DIPAR..K.LQSYMGERIQQMTAH..SSPTARCRLAQAE.....RDNRD-RATAHR 133
ERV-DC7 (Refrex-1)     55  HA.NTWW.P.TP.F.Q.AAGL...DIPTK..KDLQSHMGERTQQMTAH..SSPTARCRLAQAE.....RDNRD-RATAHR 133
ERV-DC16 (Refrex-1)    55  HA.NTWW.P.TP.F.Q.AAGL...DIPTK..KDLQTHMGERTQQMTAH..SSPTARCRLAQAE.....RDNRD-RATAHR 133
RmRV1                  62  APMGTWF.D.YFN.DKIAK----------------------IDDMEGGEWRKQA.RVSVSRNG..A...-FRT-.EMKKT 117
CmRV1                  62  APMGTWF.D.YFN.DKIAK----------------------IDDMEGGEWRKQA.RVSVSRNG..A...-FRT-.EMKKT 117
CmRV1_1                62  APMGTWF.D.YFN.DKIAK----------------------IDDMEGGEWRKQA.RVSVSRNG..A...-FRT-.EMRKT 117
CmRV1_2                62  APMGTWF.D.YFN.DKIAK----------------------IDDMEGGEWRKQA.RVSVSRNG..A...-FRT-.EMRKT 117
CERV2 (extinct)        63  AP.NTWF.D.YFN.DKIAM----------------------IDEMEGGEWRKQA.RVSLSRNG......-FRT-..MKKT 118
RhERV2-A (extinct)     62  APMGTWF.D.YFN.DKIAK----------------------IDDMEGGEWRKQA.RVSVSRNG..A...-FRT-.EMRKT 117
CmRV-trunc1            78  APIGTWF.D.YFN.DKVA.----------------------VNEMEGGEWRKQA.RVSISRNG..A...-FRT-.DMEKT 133
RmRV-trunc1            78  APIGTWF.D.YFN.DKVA.----------------------VNEMEGGEWRKQA.RVSISRNG..A...-FRT-.DMEKT 133
PNRC2-ChRV             63  AP.NTWF.D.YFN.DKIAM----------------------IDEMEGGEWRKQA.RVSLSRNG......-FRT-..MKKT 118
PNRC2-BoRV             63  AP.NTWF.D.YFN.DKIAM----------------------TDEMEGGEWRKQA.RVSLSRNG......-FRT-..MKKT 118
BoRV-7                 63  AP.NTWF.D.YFN.DKIAM----------------------IDEMEGGEWRKQA.RVSLSRNG......-FRT-..MKKT 118
CERV2 provirus (ChrY)  63  AP.NTWF.D.YFN.DKIAM----------------------INEMEGGEWRKQA..VSLSRNG......-FRT-..MKKT 118
CERV2-trunc(artificial)63  AP.NTWF.D.YFN.DKIAM----------------------IDEMEGGEWRKQA.RVSLSRNG......-FRT-..MKKT 118
GoRV-trunc1            65  AP.NTWF.D.YFN.DQIAR----------------------INEMEGGEWRKNA.RVSLSRNG..A...-FRT-..MKKT 120
                                                                  SU
                                                                                VRB
FeLV-A                 139 CGGAQDGFCAAWGCETTGEAWWKPSSSWDYITVKRGSSQD-NNCE-----------------------GKCNPLILQFTQ 194
FeLV-B                 124 ...P......V........TY.R.T.........K.VT.GIYQ.SGGGWCGPCYDKAVHSSTTGASEG.R.......... 203
FeLV-D (TY26)          134 ...YEEY..S........D.Y.Q.T....L..IT..YTKP-DPDGHT-----CYYKKGTEG-YHHWISPLSL..KIT..D 206
FeLV-D (ON-T)          134 ...YEEY..S........D.Y.Q.T....L..IT..YTKP-DPDRHT-----CYYKKGTEG-SHHWISPLSL..KIT..D 206
FeLV-D (ON-C)          132 ...YEEY..S........D.Y.Q.T....L..IT..YTKP-DPDRHT-----CYYKKGTKG-YHHWISPLSL..KIT..D 204
FeLV-D (44B)           134 ...YEEY..S........D.Y.Q.T....L..IT..YTKP-DPGRYT-----CYYKKGTEG-YHHWISPLSL..KIT..D 206
ERV-DC10               134 ...YEEY..S........D.Y.R.T....L..IT.NYTKP-DS.DDR-----VERKRKMS---RHWRDPLSL..KIT..D 204
ERV-DC14               134 ...YEEY..S........D.Y.Q.T....L..IT..YTKP-DPDGHT-----CYYKKGTEG-YHHWISPLSL..KIT..D 206
ERV-DC7 (Refrex-1)     134 ...YEEY..S........D.Y.Q.T....L..IT.NYTKP-DPDGHT-----CYYKKGTEG-YHYWESPLSL..KIT..D 206
ERV-DC16 (Refrex-1)    134 ...YEEY...........D.Y.Q.T....L..IT.NYTKP-DPDGHT-----CYYKKGTEG-YHHWESPLSL..KIT..D 206
RmRV1                  118 ..EIDAL...S.S.I..NDGE..WATKPW...MS--FV.R------------C---------TRTRYSKT..LVRIK.ED 174
CmRV1                  118 ..EIDAL...S.S.I..NDGE..WATKPW...MS--FV.R------------C---------TRTRYSKT..LVRIK.ED 174
CmRV1_1                118 ..EIDAL...S.S.I..NDGE..WATKPW...MS--FV.R------------C---------TRTRYSKT..LVRIK.ED 174
CmRV1_2                118 ..EIDAL...S.S.I..NDGE..WATKPW...MS--FV.C------------C---------TRTRYSKT..LVRIK.ED 174
CERV2 (extinct)        119 ..EIMSLY..S.S.V..NDGE..WKTQPW.V.MS--YV.P------------C---------TRTRYSAT..LIRVK.EE 175
RhERV2-A (extinct)     118 ..EIDAL...S.S.I..NDGE..WATKPW...MS--FV.R------------C---------TRTRYSKT..LVRIK.ED 174
CmRV-trunc1            134 ..DITHLY.YS.S.V.NNDGE..WATKPW...MS--FV.P------------C---------TRTRYSKN..LVRIK.ED 190
RmRV-trunc1            134 ..DITHLY.YS.S.V.NNDGE..WATKPW...MS--FV.P------------C---------TRTRYSKN..LVRIK.ED 190
PNRC2-ChRV             119 ..EIMSLY..S.S.V..NDGE..WKTQPW.L.MS--YV.P------------C---------TRTRYSAT..LIRVK.EE 175
PNRC2-BoRV             119 ..EIMSLY..S.S.V..NDGE..WKTQPW.L.MS--YV.P------------C---------TRTRYSAT..LIRVK.EE 175
BoRV-7                 119 ..EIMSLY..S.S.V..NDGE..WETQPW.L.MS--YV.P------------C---------TRTRYSAT..LIRVK.EE 175
CERV2 provirus (ChrY)  119 ..EIMSLY..S.S.V..NDGE..WKTQLW.V.MS--YV.P------------C---------TRTRYSAT..LICVK.EE 175
CERV2-trunc(artificial)119 ..EIMSLY..S.S.V..NDGE..WKTQPW.V.MS--YV.P------------C---------TRTRYSAT..LIRVK.EE 175
GoRV-trunc1            121 ..EITSLY..S.S.V..NDGES.WKTQPW.V.MT--YVRP------------C---------TRTRYSAT..LIRIK.EE 177
```

```
                                                                   SU
                                                                                              PRR
FeLV-A                 195 KGKQA-SWDGPKMWGLRLYRTGYDPIALFTVSRQVSTITPPQAMGPNLVLPDQKPPSRQS--QTGSK------------- 258
FeLV-B                 204 ..R.T-......S.......S........S.....M........................--.IE.R------------- 267
FeLV-D (TY26)          207 S..R.LG.QTGYT....W.LP.K.RGIVLKIKLKID...--.TV......A...A.VQLAIPVQPPR------------- 271
FeLV-D (ON-T)          207 S..R.LG.QTGYT....W.LP.K.RGIVLKIKLKID...--.PV......A...A.VQLAIPVQPPR------------- 271
FeLV-D (ON-C)          205 S..R.LG.QTGYT....W.LP.K.RGVVLKIKLKID...--.PV......A...A.VQLAIPVQPPR------------- 269
FeLV-D (44B)           207 S..R.LG.QTGYT....W.LP.K.RGIVLKIKLKID...--.TV......A...A.VQLAIPVQPPR------------- 271
ERV-DC10               205 S..R.LG.QTGYT....W.LP.K.RGIILKIKLKID...--.TV......A...A.VQLAIPAQPPR------------- 269
ERV-DC14               207 S..R.LG.QTGYT....W.LP.K.RGIVLKIKLKID...--.TV......A...A.VQLAIPVQPPR------------- 271
ERV-DC7 (Refrex-1)     207 S..R.LG.QTGYT....W.LP.K.RGVVLKIKLKID...--.TV..T* 251
ERV-DC16 (Refrex-1)    207 S..R.LG.QTGYT....W.LP.K.RGVILKIKLKID...--.PV......A...V.VQLAIPVQPP.------------- 271
RmRV1                  175 AA.SDNR.ISGLI...Y..QKPL-YGIPIQIKLI.NP..A.V.V...Q..SETRK.LVPAPREPQPR------------- 240
CmRV1                  175 AA.SDNR.ISGLI...Y..QKPL-YGIPIQIKLI.NP..A.V.V...Q..SETRK.LVPAPREPQPR------------- 240
CmRV1_1                175 AA.SDNR.ISGLI...Y..QKPL-YGIPIQIKLI.NP..A.V.V...Q..SETRK.LVPAPREPQPR------------- 240
CmRV1_2                175 AA.SDNR.ISGLI...Y..QKPL-YGIPIQIKLI.NP..A.V.V...Q..SKTRK.LVPAPREPQPR------------- 240
CERV2 (extinct)        176 AA.TDPR.TTGLI...N..Q.PA-SGLPIQIRLL.NPVSASVPV...P..TGRA.SQSG.RQKVPTTGAQSPPAVKPPTS 254
RhERV2-A (extinct)     175 AA.SDNR.ISGLI...Y..QKPL-YGIPIQIKLI.NP..A.V.V...Q..SETRK.LVPAPREPQPR------------- 240
CmRV-trunc1            191 AA.SDN..ITGLI...Y..QKPL-YGIPIQIRLL.DPDIA.V.V.L.Q..SEE.K.PVPIPEKPQP.------------- 256
RmRV-trunc1            191 AA.SDN..ITGLI...Y..QKPL-YGIPIQIRLL.DPDIA.V.V...Q..SEE.K.PVPIPEKPQP.------------- 256
PNRC2-ChRV             176 AA.TDPR.TTGLI...N..Q.PA-AGLPIQIRLL.NPVSASVPV...P..TGRA.SQ* 231
PNRC2-BoRV             176 AA.TDPR.TTGLI...N..Q.PA-AGLPIQIRLL.NPVSASVPV...P..TGRA.SQ* 231
BoRV-7                 176 AA.TDPR.TTGLI...N..Q.PA-SGLPIQIRLL.NPVSASVPV...P..TGRA.S* 230
CERV2 provirus (ChrY)  176 AA.TDPR.TTGLI* 188
CERV2-trunc(artificial)176 AA.TDPR.TTGLI* 188
GoRV-trunc1            178 AA.TDSR.TSGLI...N..Q.PA-FGVPIQIKLI.NPASA.VPI...Q..TGKA...SR.RQKVPTA------------- 243
                                                                   SU
                                                             PRR
FeLV-A                 259 ---------------------VATQRPQTNESAP------RSVAP-----T---------TVGPKRIGTGDRLINLVQGT 297
FeLV-B                 268 ---------------------.TPHHS.G.GGT.GITLVNA.I..LSTPV.---------PAS................. 317
FeLV-D (TY26)          272 ---------------------AP..-------T.GINPVNSTLS.SLGYP.-----------P.S.P......L.....V 312
FeLV-D (ON-T)          272 ---------------------AP..-------T.GINPVNSTLS.SLGYP.-----------P.S.P......L.....V 312
FeLV-D (ON-C)          270 ---------------------AP..-------T.GINPVNSTLS.SLGYP.-----------P.S.P......L.....V 310
FeLV-D (44B)           272 ---------------------AP..-------T.GINPVNSTLS.SLGYP.P----------P.S.P......L....RV 313
ERV-DC10               270 ---------------------AP..-------T.RVNPVNSTPS.SLGYPAPAPGPRPPYPTS.S.P......L.....V 321
ERV-DC14               272 ---------------------AP..-------T.GINPVNSTLS.SLGYP.-----------P.S.P......L.....V 312
ERV-DC16 (Refrex-1)    272 ---------------------AP..-------T.RVSPANSTLS.NLGYS.-----------S.S* 297
RmRV1                  241 ---------------------APKS------TS.PLISTSSKYT.SAQTV.R----------..LNL.I....L..IK.S 283
CmRV1                  241 ---------------------APKS------TS.PLISTSSKYT.SAQTV.R----------..LNL.I....L..IK.S 283
CmRV1_1                241 ---------------------APKS------TS.PLISTSSKYT.SAQTV.R----------..LNL.I....L..IK.S 283
CmRV1_2                241 ---------------------APKS------TS.PLISTSSKYT.SAQNV.R----------..LDL.I....L..IK.S 283
CERV2 (extinct)        255 PSQSGDSVITGKAPPQSGSPPSPPP------TS.SALPGTTRLP.DPET--------------------SN..F..IR.A 308
RhERV2-A (extinct)     241 ---------------------APKS------TS.PLISTSSKYT.SAQNV.R----------..LDL.I....L..IK.S 283
CmRV-trunc1            257 ---------------------APQS------TS.PLISASSEYTSSAQNV.R* 281
RmRV-trunc1            257 ---------------------APQS------TS.PLISTSSEYTSSAQNV.R* 281
GoRV-trunc1            244 ---------------------APPT------.S.SALPGTTPLP.DPET--------------------SN..FH.IR.A 276

                                                                   SU
FeLV-A                 298 YLALNATDPNKTKDCWLCLVSRPPYYEGIAILGNYSNQTNPPPSCLSIPQHKLTISEVSGQGLCIGTVPKTHQALCNKTQ 377
FeLV-B                 318 ..........R..................................................................E.. 397
FeLV-D (TY26)          313 ..T..L.A.SQ.Q......TAK....Q.V..I..FT.H..A.LR.STT.R.G..LT..T.H.....KI.PS..N..SQ.I 392
FeLV-D (ON-T)          313 ..T..L.A.SQ.Q......TAK....Q.V..I..FT.H..A.LR.STT.R.G..LT..T.H.....KI.PS..N..SQ.I 392
FeLV-D (ON-C)          311 ..T..L.A..Q.Q......TAK....Q.V..I..FT.H..A.LR.STT.R.G..LT..T.H.....KI.PS..N..SQ.I 390
FeLV-D (44B)           314 ..T..L.A..Q.Q......TAK....Q.V..I..FT.H..A.LR.STT.R.G..LT..T.H.....KI.PS..N..SQ.I 393
ERV-DC10               322 ..T..L.T..Q.Q......TAK....Q.V..I..FT.H..A.LR.STT.R.G..LT..T.H.....KI.PS..N..SQ.I 401
ERV-DC14               313 ..T..L.A..Q.Q......TAK....Q.V..I..FT.H..A.LR.STT.R.G..LT..T.H.....KI.PS..N..SQ.I 392
RmRV1                  284 .F...Q.K.EF.SS.....ATG........STN.FT.SA.-.TG.AWEQ.R...LA....S.T...Q..PS..H...V.L 362
CmRV1                  284 .F...Q.K.EF.SS.....ATG........STN.FT.SA.-.TG.AWEQ.R...LA....S.T...Q..PS..H...V.L 362
CmRV1_1                284 .F...Q.K.EF.SS.....ATG........STN.FT.SA.-.TG.AWEQ.R...LA....S.T...Q..PS..H...V.L 362
CmRV1_2                284 .F...Q.K.EF.SS.....ATG........STN.FT.SA.-.TG.AWEQ.R...LA....S.T...Q..PS..H...V.L 362
CERV2 (extinct)        309 .....Q.R.ES.TS.....A.S........SIS.FT.S.S-HSG.AWDQHK...LA....S.T...R..PS..H...R.L 387
RhERV2-A (extinct)     284 .F...Q.K.EF.SS.....ATG........STN.FT.SA.-.TG.AWEQ.R...LA....S.T...Q..PS..H...V.L 362
GoRV-trunc1            277 .....Q.K.ES.AS.....ATG........SVS.FT.S..-HSG.AWERHK...LS....S.T...K..PS..H...V.L 355

                                                          SU                                             TM
FeLV-A                 378 QG-HTGAHYLAAPNGTYWACNTGLTPCISMAV--LNWTSDFCVLIELWPRVTYHQPEYVYTHFAKA--VRFRREPISLTV 452
FeLV-B                 398 ..-.............................--................................--A........... 472
FeLV-D (TY26)          393 PSVGQ.PY..T.......V..........LQ.--..N.A.Y.I......KIF..DS..I.G.YEPG--G......V.... 468
FeLV-D (ON-T)          393 PSVGQ.PY..T.......V..........LQ.--F.N.A.Y.I.......IF..DS..I.G.YEPG--G......V.... 468
FeLV-D (ON-C)          391 PSVGQ.PY..T.......V..........LQ.--..N.A.Y.I......KIF..DS..I.G.YEPG--G......V.... 466
FeLV-D (44B)           394 PSVGQ.PY..T.......V..........LQ.--..N.A.Y.I......KIF..DS..I.G.YEPG--G......V.... 469
ERV-DC10               402 PSVGQ.PY..T.......V..........LQI--..D.A.Y.I......KIF..DS..I.G.YELG--G....D.V.... 477
ERV-DC14               393 PSVGQ.PY..T.......V..........LQ.--..N.A.Y.I......KIF..DS..I.G.YEPG--G......V.... 468
RmRV1                  363 TV-PSSN...VPSETDW............T..--FSSGTHY...VQVV...Y..SGDSFDLRYEQKTHT.PK.......L 439
CmRV1                  363 TV-PSSN...VPSETDW............T..--FSSGTHY...VQVV...Y..SGDSFDLRYEQKTHT.PK.......L 439
CmRV1_1                363 TV-PSSN...VPSETDW............T..--FSSGTHY...VQVV...Y..SGDSFDLRYEQKTHT.PK.......L 439
CmRV1_2                363 TV-PSSN...VPSETDW..........V.T..--FSSGTHY...VQVV...Y..SGDSFDLRYEQKTHT.PK.......L 439
CERV2 (extinct)        388 AV-PRTS...IPSGPDW...K......V.T..--F.NSE.Y.I.VQVV...Y.QTG.SFESQ.EQKYLT.MK...V...L 464
RhERV2-A (extinct)     363 TV-PSSN...VPSETDW..........V.T..--FSSGTHY...VQVV...Y..SGDSFDLRYEQKTHT.PK.......L 439
GoRV-trunc1            356 TV-PSTS...IPSGTDW...D......LPL.YPQ.SS.T-----VKIIVHCY---------------------------- 401
```

```
                                                            TM
FeLV-A                 453 ALMLGGLTVGGIAAGVGTGTKALLETA-QFRQLQMAMHTDIQALEESISALEK-SLTSLSE-VVLQNRRGLDILFLQEGG 529
FeLV-B                 473 .......................I...-.........................-.......-.................. 549
FeLV-D (TY26)          469 ..L.....M.SL...I....A..I..N-..K...I...S.............R-.......-..........L....... 545
FeLV-D (ON-T)          469 ..L.....M.SL...I....A..I..N-..K...I...S.............R-.......-..........L....... 545
FeLV-D (ON-C)          467 ..L.....M.SL...I....A..I..N-..K...I...S..............-.......-.................. 543
FeLV-D (44B)           470 ..L.....M.SL...I....A..I..N-..K...I...S.............R-.......-..........L....... 546
ERV-DC10               478 ..L.....M.SL...I....A..I..N-..K...I...S.............R-.......-..........L....... 554
ERV-DC14               469 ..L.....M.SL...I....A..I..N-..K...I...S.............R-.......-..........L....... 545
RmRV1                  440 .V...----I.V........A..VHGNHHLQ..RV.IDE.LR.I.Q..TK..E-.......-.........E.V..K... 513
CmRV1                  440 .V...----I.V........A..VHGNHHLQ..RV.IDE.LR.I.Q..TK..E-.......-.........E.V..K... 513
CmRV1_1                440 .V...----I.V........A..VHGNHHLQ..RV.IDE.LR.I.Q..TK..E-.......-.........E.V..K... 513
CmRV1_2                440 .V...----I.V........A..VHGNHHLQ..RV.IDE.LR.I.Q..TK..E-.......-.........E.V..K... 513
CERV2 (extinct)        465 .V...----L.V........A..VRGSYHLQ..RA.VDE.LR.I.H..TK..E-.......-...........I..K... 538
RhERV2-A (extinct)     440 .V...----I.V........A..VHGNHHLQ..RV.IDE.LR.I.Q..TK..E-.......-.........E.V..K... 513
GoRV-trunc1            402 -----------------------------------------RLCP.F.TK..NL.SS...KNIS.EG.EN.---.PSPSL 437

                                                            TM
FeLV-A                 530 LCAALKEECCFYADHTGLVRDNMAKLRERLKQRQQLFDSQQGWFEGWFNRSPWFTTLISSIMGPLLILLLILLFGPCILN 609
FeLV-B                 550 .................................................K.............................. 629
FeLV-D (TY26)          546 .....................S...........................................I.............. 625
FeLV-D (ON-T)          546 .................I...S............K..E.........Y.K.......V..L....IP.....M....... 625
FeLV-D (ON-C)          544 ....................E....................................L...I...I.............. 623
FeLV-D (44B)           547 .....................S.......................R...................I.......L...... 626
ERV-DC10               555 .................I...S............K..E.........Y.K.......V..L....IL.....V....... 634
ERV-DC14               546 .................I...S............K..E.........Y.K.......V..L....IL.....M....... 625
RmRV1                  514 .......Q.......S.V.K.S........DK.KKERE...N...N.Y.Q...LS....T.L...IL.M...T....... 593
CmRV1                  514 .......Q.......S.V.K.S........DK.KKERE...N...N.Y.Q...LS....T.L...IL.M...T....... 593
CmRV1_1                514 .......Q.......S.V.K.S........DK.KKERE...N...N.Y.Q...LS....T.L...IL.T...T....... 593
CmRV1_2                514 .......Q.......S.V.K.S........DK.KKERE...N...N.Y.Q...LS....T.L...IL.T...T....... 593
CERV2 (extinct)        539 .......Q.......S.V.K.S.....K..DD..KERE...S...T.Y.Q....S....T.L...IL.M...T.....F. 618
RhERV2-A (extinct)     514 .......Q.......S.V.K.S........DK.KKERE...N...N.Y.Q...LS....T.L...IL.T...T....... 593
GoRV-trunc1                ..* 439

                               TM               R peptide
FeLV-A                 610 RLVQFVKDRISVVQALILTQQYQQIKQYDPDRP*     642
FeLV-B                 630 .................................*     662
FeLV-D (TY26)          626 .........M..I..................Q.*     658
FeLV-D (ON-T)          626 ......R.........V..............Q.*     658
FeLV-D (ON-C)          624 ................V..............Q.*     656
FeLV-D (44B)           627 ...............................Q.*     659
ERV-DC10               635 .....IRE.L..I...V.....H.LR.F.AE..DTIE* 671
ERV-DC14               626 .....IRE.L..I...V.....H.LR.F.AE..DTIE* 662
RmRV1                  594 ..LTLI.N.LNI.H.MV......TLRTEEEAQD*     626
CmRV1                  594 ..LTLI.N.LNI.H.MV......TLRTEEEAQD*     626
CmRV1_1                594 ..LTLI.N.LNI.H.MV......TLRTEEEAQD*     626
CmRV1_2                594 ..LTLI.N.LNI.H.MV......TLRTEEEAQD*     626
CERV2 (extinct)        619 ..LTLI...LNI.H.M.......AV.TDEETQD*     651
RhERV2-A (extinct)     594 ..LTLI.N.LNI.H.MV......TLRTEEEAQD*     626
```

Figure S7. Alignment of the amino acid sequences of primate ERV Envs, feline ERV Envs, and feline gammaretrovirus Envs. FeLV-A, FeLV-B, FeLV-D (TY26, ON-T, ON-C, and 44B), ERV-DC10, ERV-DC14, Refrex-1 (ERV-DC7 and ERV-DC16), CmRV1, CmRV1_1, CmRV1_2, RmRV1, CERV2, RhERV2-A, GoRV-trunc1, CmRV-trunc1, RmRV-trunc1, PNRC2-ChRV, PNRC2-BoRV, BoRV-7, CERV2 provirus (ChrY), and CERV2-trunc Envs are shown using single-letter amino acid codes. Threonine (T) at position 82 in PNRC2-BoRV is highlighted in red. Dots, conserved amino acid residues; hyphens, gaps in the amino acid sequence; asterisks, stop codons. SU, surface unit; TM, transmembrane domain; VR, variable region; PRR, proline-rich region.

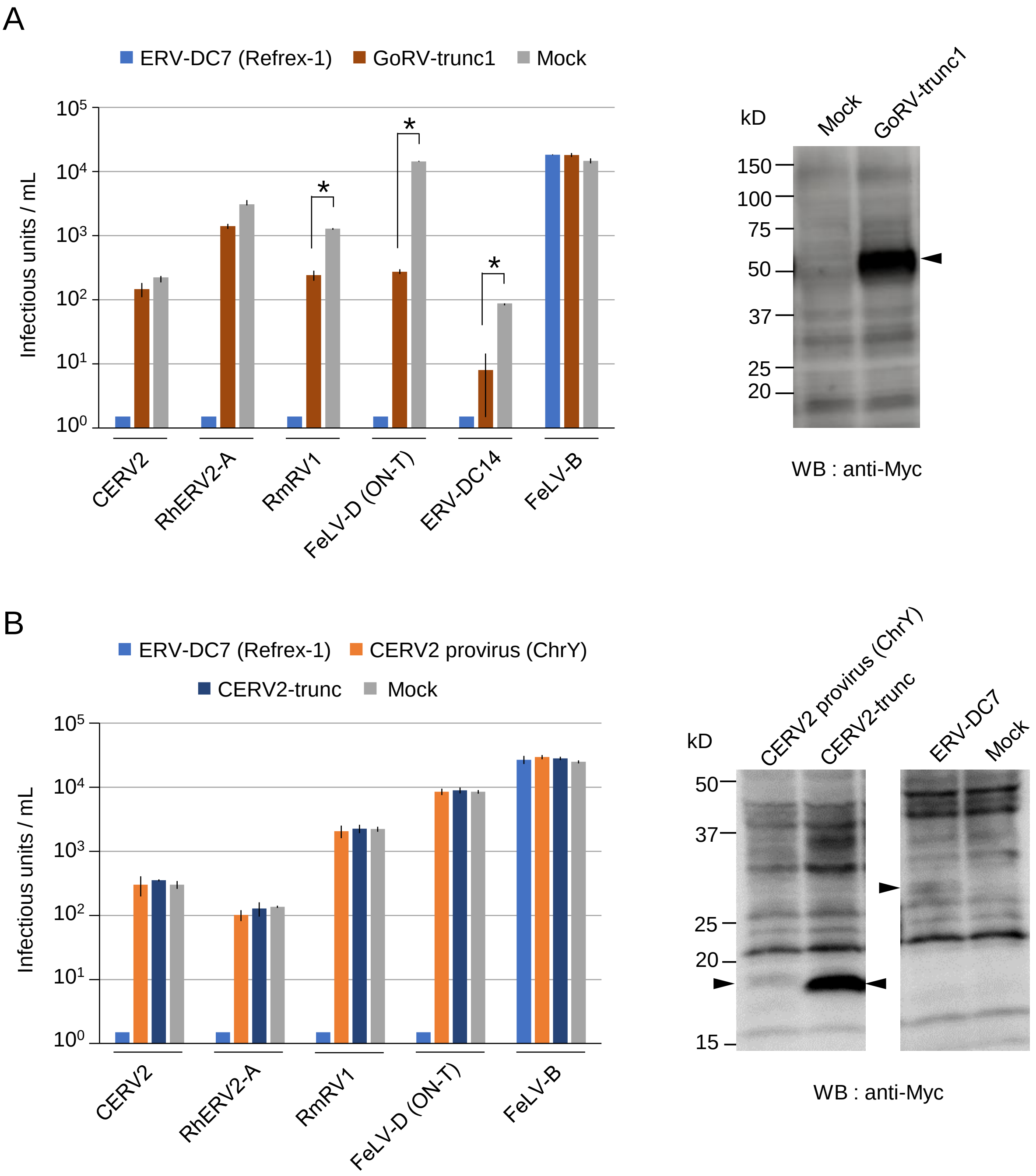
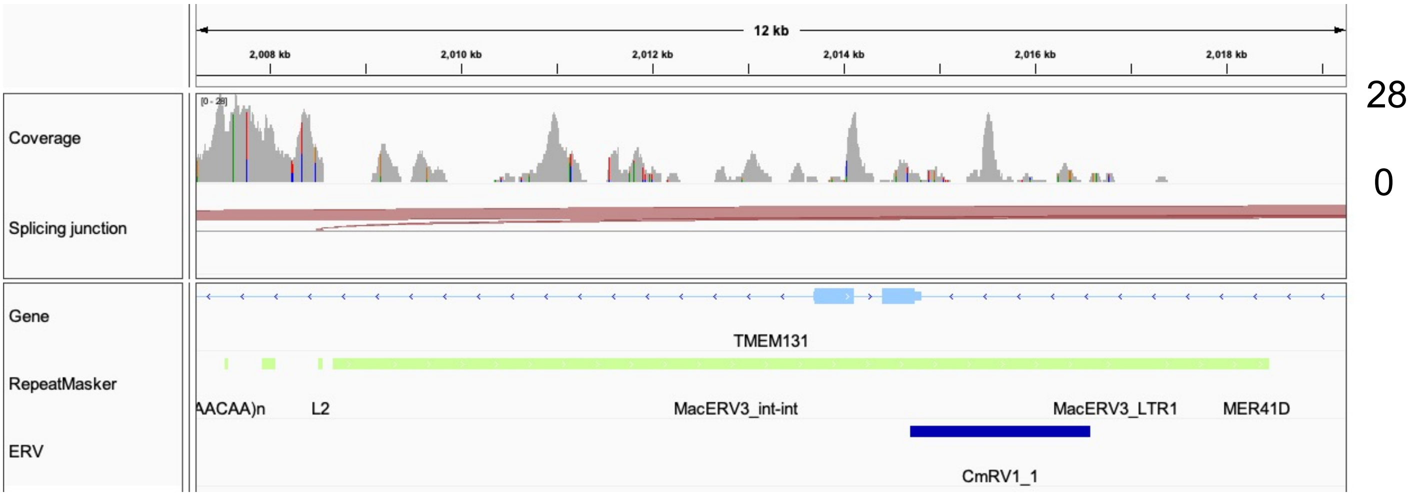


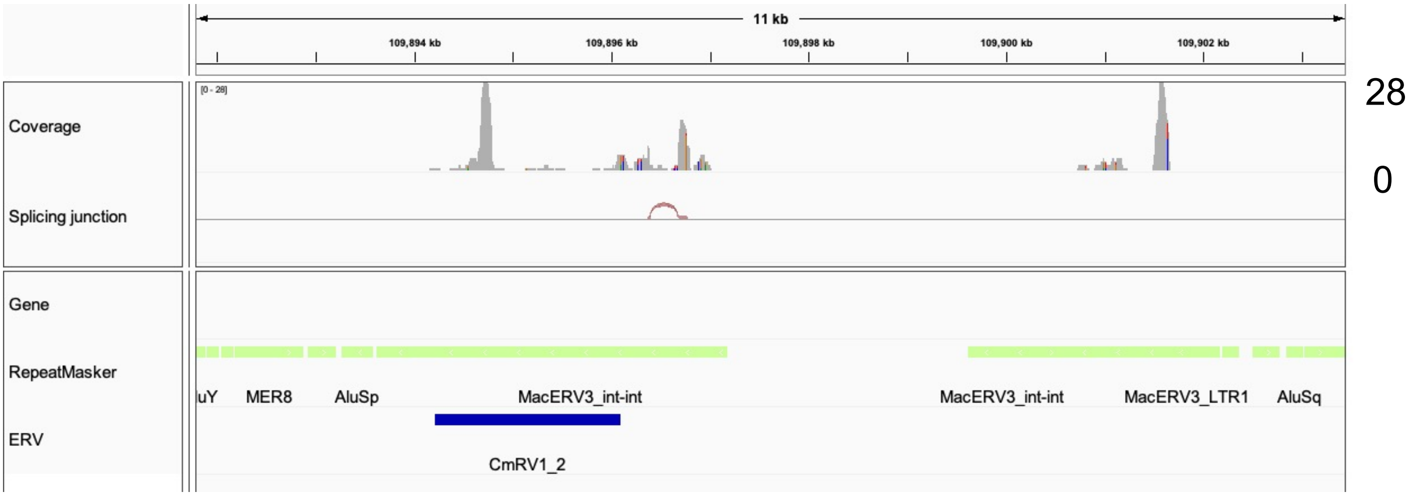
Figure S8. Inhibitory effect of GoRV-trunc1, CERV2 provirus (ChrY), and CERV2-trunc against retroviral infection. (A) HEK293T cells were transfected with expression vectors encoding GoRV-trunc1, ERV-DC7 (Refrex-1), or empty vector (Mock) (A, left side). **(B)** HEK293T cells were transfected with expression vectors encoding ERV-DC7 (Refrex-1), CERV2 provirus (ChrY; Chromosome Y), CERV2-trunc (constructed from CERV2 Env), or empty vector (Mock) (B, left side); cells were then infected with Env-pseudotyped viruses indicated on the x-axis. The y-axis indicates infectious units per milliliter of supernatant. Infection titers are means of three infection experiments. Differences between groups were assessed using Student's t-test (*, $p < 0.01$). Bars represent standard deviations. Western blotting (WB) with anti-c-Myc antibody of cell lysates from HEK293T cells transfected with the indicated plasmids. Arrows shown in the figure indicate specific bands (A and B, right side).

Crab-eating macaque (Thymus: SRR1758961)

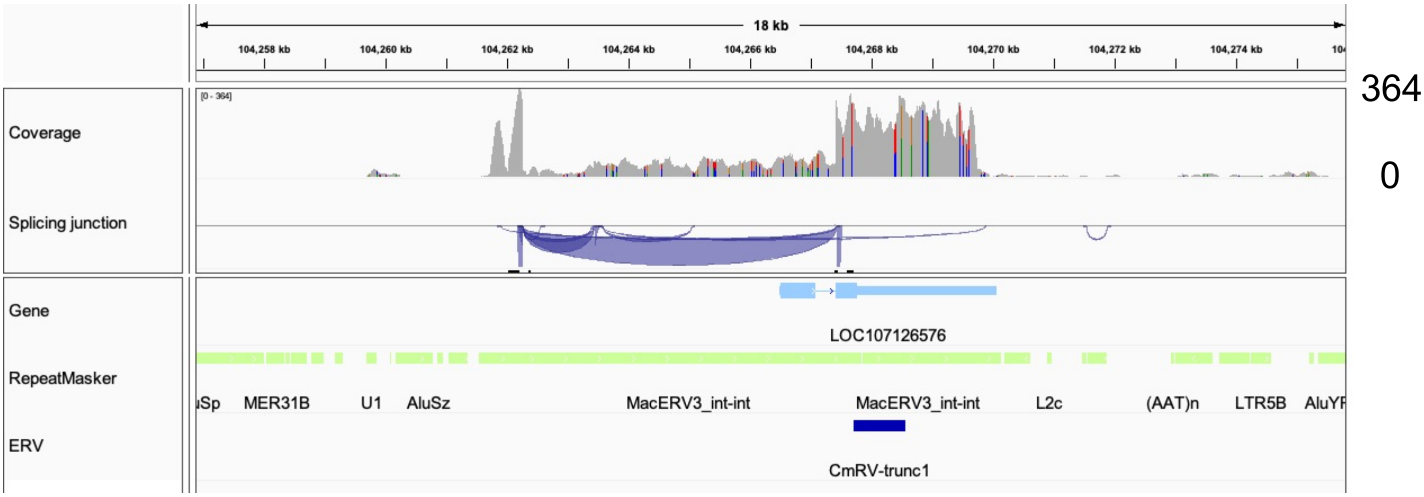
CmRV1_1 [Chr 13: 2,007,224-2,019,238]



CmRV1_2 [Chr 14: 109,891,789-109,903,426]

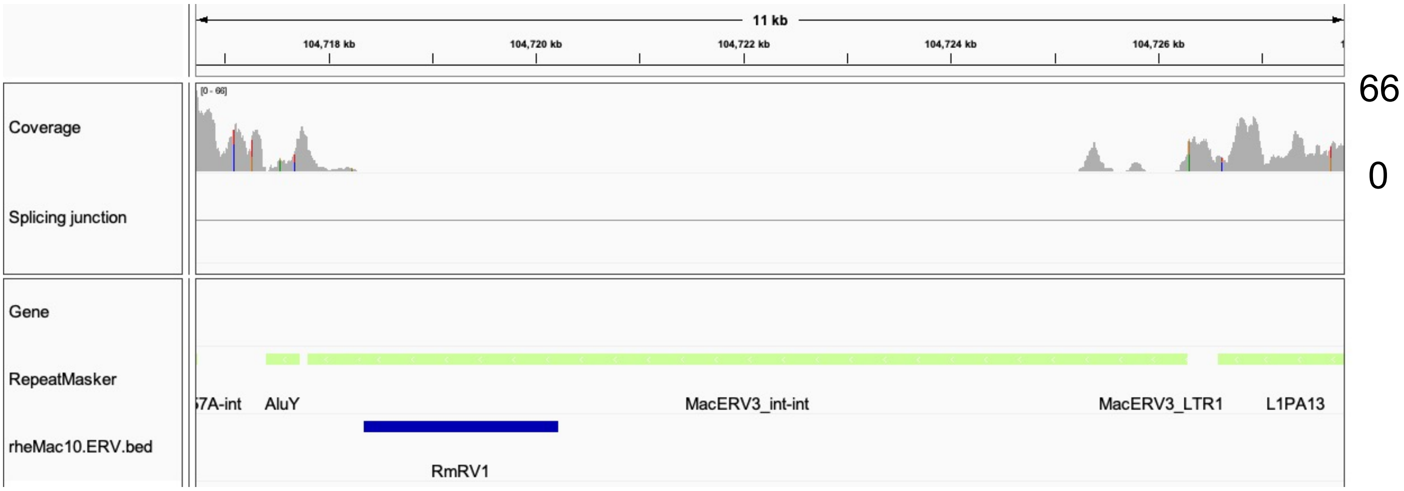


CmRV-trunc1 [Chr 1: 104,256,885-104,275,785]

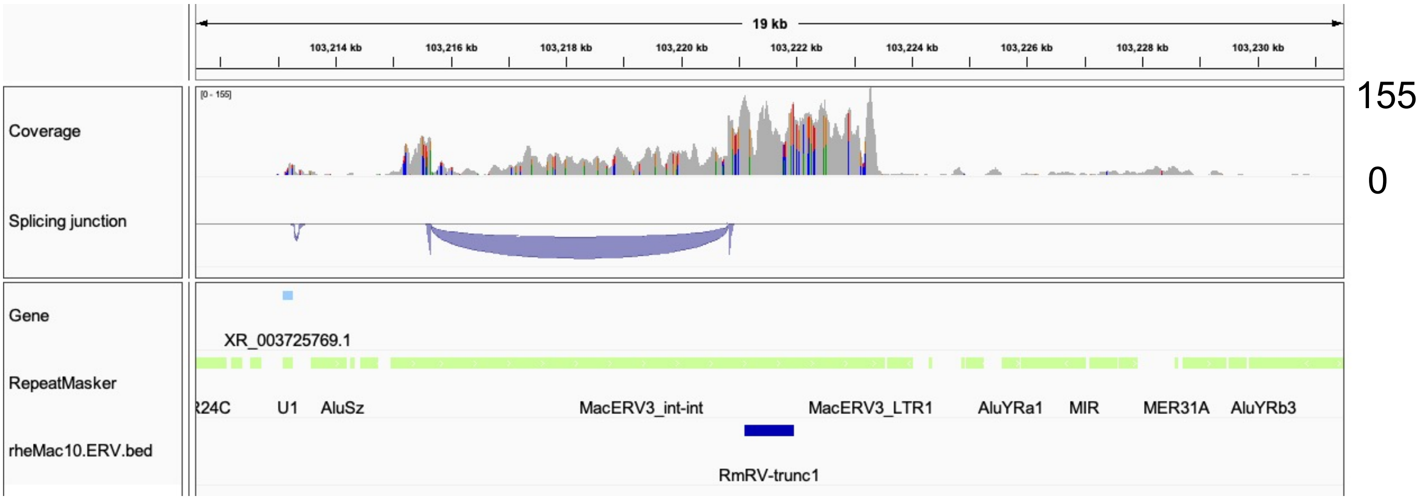


Rhesus macaque (Thymus: SRR1602572)

RmRV1 [Chr 11:104,716,721-104,727,782]

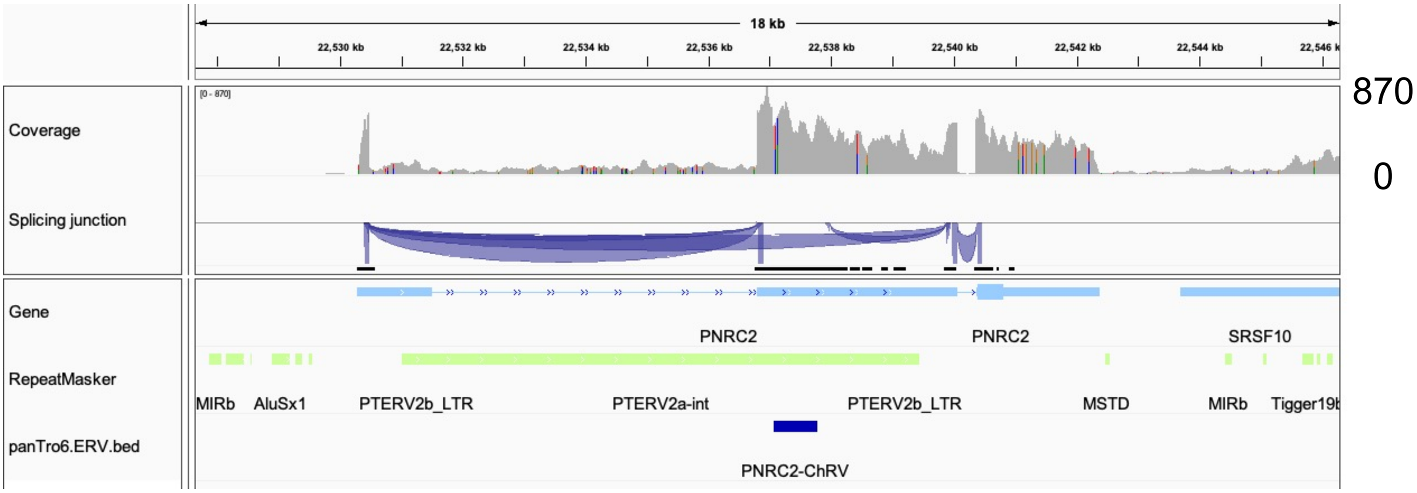


RmRV-trunc1 [Chr 1: 103,211,574-103,231,480]



Chimpanzee (Thymus: SRR1758933)

PNRC2-ChRV [Chr 1: 22,527,643-22,546,267]



CERV2 provirus (ChrY) [Chr Y: 20,220,009-20,230,217]

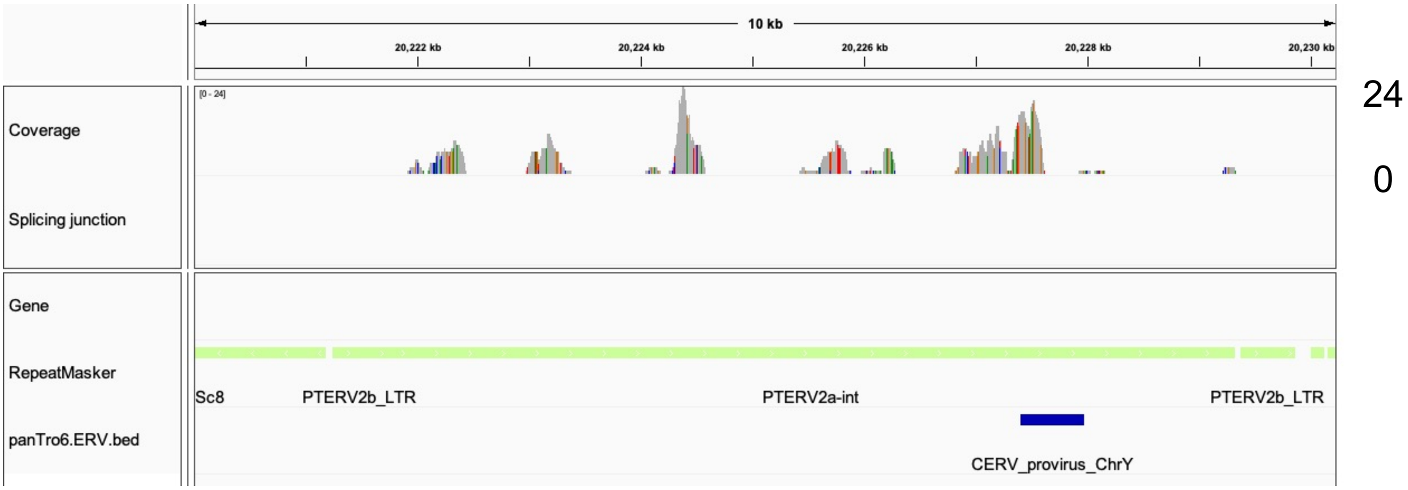


Figure S9. The mapped pattern of RNA-seq reads to genomic regions encoding truncated [CmRV-trunc1, RmRV-trunc1, PNRC2-ChRV, and CERV2 provirus (ChY)] and full-length Env (CmRV1_1 and RmRV1_2) proteins. Each panel shows mapped patterns in the thymus data of crab-eating macaque (SRR1758961), rhesus macaque (SRR1602572), and chimpanzee (SRR1758933). The annotations of host genes (light blue), repetitive elements defined by RepeatMasker (light green), and primate *env* genes (blue) are shown. The number shown to the right of the coverage column indicates the range of RNA-seq read counts. Expression of truncated *env* genes (CmRV-trunc1, RmRV-trunc1, and PNRC2-ChRV) is found as spliced transcripts.


A
Primate ERVs
RmRV1
CmRV1
RhERV2-A*
RmRV-trunc1
CmERV-trunc1
PNRC2-BoRV
PNRC2-ChRV
BoRV-7
GoRV-trunc1
BaEV
ERV-DC related viruses in feline
ERV-DC8
ERV-DC14
RD114
ERV-DC10
ERV-DC7
ERV-DC16
enFeLV
FeLV-D (ON-T)
FeLV-A
M-MLV
F-MLV
KoRV
GALV
PERV-C
RfRV
REV
MMTV
JSRV
ALV
Gammaretrovirus
0.3
B
Primate ERVs
RmRV1
CmRV1
RhERV2-A*
RmRV-trunc1
CmRV-trunc1
BaEV
PNRC2-BoRV
PNRC2-ChRV
CERV2 provirus (ChrY)
BoRV-7
GoRV-trunc1
ERV-DC related viruses in feline
RD114
ERV-DC10
ERV-DC8
ERV-DC14
ERV-DC7
ERV-DC16
M-MLV
F-MLV
enFeLV
FeLV-A
KoRV
GALV
PERV-C
RfRV
REV
MMTV
JSRV
ALV
Gammaretrovirus
0.5

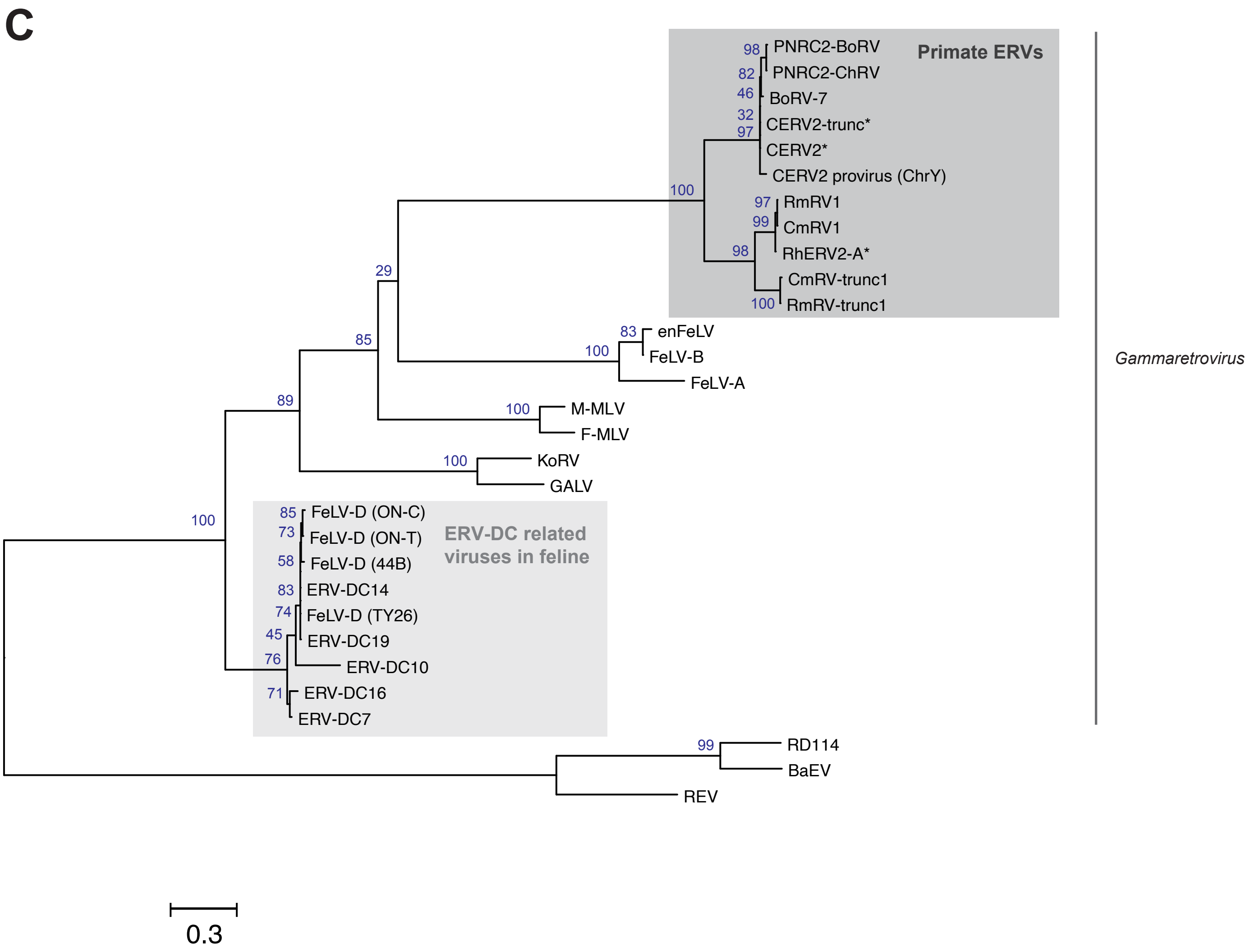
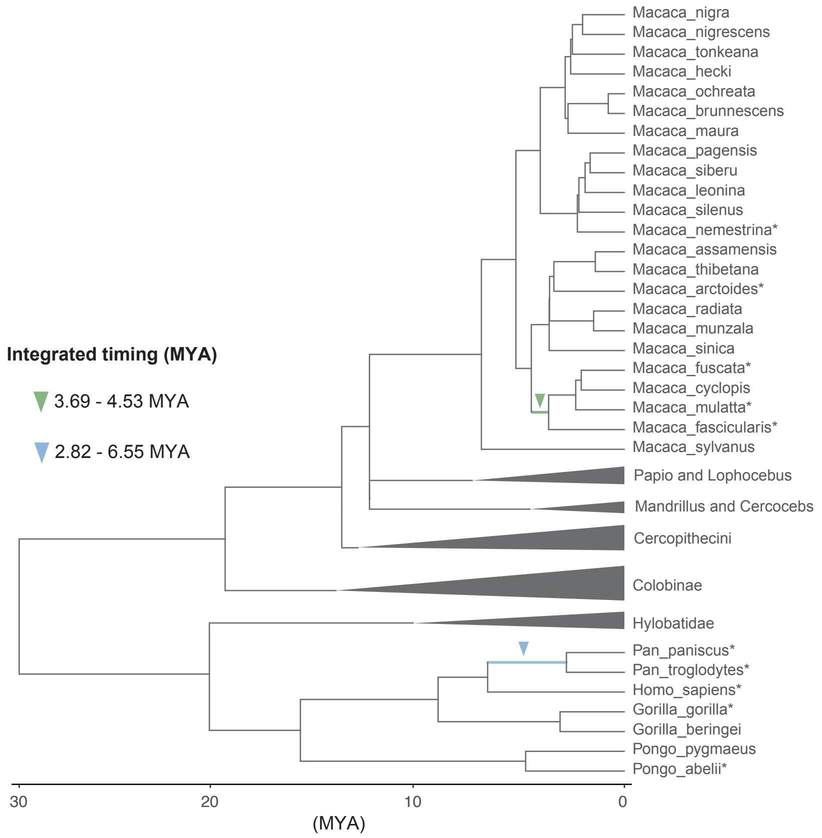

C
Primate ERVs
98 PNRC2-BoRV
82 PNRC2-ChRV
46 BoRV-7
32 CERV2-trunc*
97 CERV2*
CERV2 provirus (ChrY)
100
97 RmRV1
99 CmRV1
98 RhERV2-A*
CmRV-trunc1
100 RmRV-trunc1
29
83 enFeLV
100 FeLV-B
FeLV-A
85
100 M-MLV
F-MLV
89
100 KoRV
GALV
100
ERV-DC related viruses in feline
85 FeLV-D (ON-C)
73 FeLV-D (ON-T)
58 FeLV-D (44B)
83 ERV-DC14
74 FeLV-D (TY26)
45 ERV-DC19
76 ERV-DC10
71 ERV-DC16
ERV-DC7
99 RD114
BaEV
REV
Gammaretrovirus
0.3

Figure S10. Phylogenetic analyses using gammaretroviruses. Phylogenetic tree using amino acid sequences of the *gag* gene **(A)**, the *pol* gene **(B)**, and the *env* gene **(C)**. The phylogenetic trees were constructed based on the maximum likelihood method for the *gag, pol* and the neighbor-joining method for the *env* genes. Scale bars indicate genetic distance (substitutions per site). Blue labels on branches indicate bootstrap support values (%) with 1,000 replicates. Colored boxes highlight CERV-related primate ERVs (dark gray) and ERV-DC related viruses in feline genomes (light gray). The asterisks indicate artificially synthesized sequences.

A

ERV	Host	Integrated age (MYA)	Ortholog (Species)
CmRV1	Crab-eating macaque *(Macaca fascicularis)*	unknown	N.D.
CmRV1_1	Crab-eating macaque *(Macaca fascicularis)*	unknown	N.D.
CmRV1_2	Crab-eating macaque *(Macaca fascicularis)*	3.69 - 4.53	chr14 : 107027464-107029344 (Rhesus macaque)
CmRV-trunc1	Crab-eating macaque *(Macaca fascicularis)*	3.69 - 4.53	RmRV-trunc1 (Rhesus macaque)
RmRV1	Rhesus macaque *(Macaca mulatta)*	unknown	N.D.
RmRV-trunc1	Rhesus macaque *(Macaca mulatta)*	3.69 - 4.53	CmRV-trunc1 (Crab-eating macaque)
PNRC2-BoRV	Bonobo *(Pan paniscus)*	2.82 - 6.55	PNRC2-ChRV (Chimpanzee)
BoRV-7	Bonobo *(Pan paniscus)*	unknown	N.D.
PNRC2-ChRV	Chimpanzee *(Pan troglodytes)*	2.82 - 6.55	PNRC2-BoRV (Bonobo)
CERV2 provirus (ChrY)	Chimpanzee *(Pan troglodytes)*	unknown	N.D.
GoRV-trunc1	Gorilla *(Gorilla gorilla)*	unknown	N.D.

B



(*: the species genome has been reported.)

Figure S11. Integration timings and orthologous relationships of CERV-related ERVs. (A) The integration timings of primate ERVs were estimated by the ortholog-dating method. MYA, million years ago; N.D., orthologs of the *env* gene were not detected; unknown, integration timing is unknown. All sequences represent Env proteins from primates including crab-eating macaque (CmRV1_1, CmRV1_2, and CmRV-trunc1), rhesus macaque (RmRV1 and RmRV-trunc1), bonobo (PNRC2-BoRV and BoRV-7), chimpanzee (PNRC2-ChRV and CERV2 provirus ChrY), and gorilla (GoRV-trunc1). Chr, chromosome. **(B)** Evolutionary history of primate species with the integration timings of CERV-related ERVs. The arrowheads indicate integration timings of the ancestor of CmRV-trunc1 and RmRV-trunc1 and that of CmRV1_2 (green), and that of PNRC2-ChRV and PNRC2-BoRV (blue).

```
                            5' flanking
CmRV-trunc1     1      CCTGTTTTTACATTCGTATTTTTGTTTCTGGGAAATATTCTTGACATTATCTTCAACTTAATTTTTGTTCTTGCTTCTAC 80
RmRV-trunc1     1      ...............T................................................................ 80

CmRV-trunc1     81     TATTTTAGTTTAATTTCCAAGTATTTTGATATTGAACTTTCAAGAGAACAGACATCGCTGCCAGCCTTGACGTCTAATGC 160
RmRV-trunc1     81     ........................................................A....................... 160

CmRV-trunc1     161    TTTGGGGGCGAGGGGGAAATGTTTTTGTTTTTTTAATACTCAGTCACCTCTTGTCACCCAGGGTGGAGTGCATTTTAACT 240
RmRV-trunc1     161    ...................................G............................................ 240

CmRV-trunc1     241    TTAATGCGTTTTAACGAAGGTATGCACACAGTGCCACATCACAAAGATAACCAGGTTTGAAAGGAATAGGTTTCCTTTGT 320
RmRV-trunc1     241    ................................................................................ 320

CmRV-trunc1     321    AATCCTAAACGTTCTAGTCTGCTCATTGAAAGCCATTATTTGAAGAAGGGTGCACGGGCTTCCAGCTGGCCGCCAAAGGG 400
RmRV-trunc1     321    ................................................................................ 400

CmRV-trunc1     401    GACCTCCACAGGACACAGATTGAGTACCCGCTTCTTTGCGGGAGACGAAGGGCGGGGACAGGAGGGAAAGAAGGGAGAGG 480
RmRV-trunc1     401    ................................................................................ 480

CmRV-trunc1     481    TATAAGTCACTTCCTCCCGGCTCCGGCGGCGGGTTGGTCCGCTGAGTTGCGGAGGAACAGACGGGGACCCGGAATGCGCT 560
RmRV-trunc1     481    ................................................................................ 560

CmRV-trunc1     561    GTCGGTGACATCACGGAGGGGGCGACTTCTATGTAGATGAGGCAGCGCAGGGGCTGCCGCTTCGCCGCCTGCTGCTTCGC 640
RmRV-trunc1     561    ................................................................................ 640

CmRV-trunc1     641    CACGAAGGGGTTCCCGTGCTGTGGGAGCGAGTCTGGGACCACTGGTCGGAACTGAGAATTCCAGCTGTGTGTCAGGGCTA 720
RmRV-trunc1     641    ................................................................................ 720

CmRV-trunc1     721    GGAGGGCTCCGGGATGCGCGCGAGGCAAGTGACCGTGTGTGTAAAAGGGTGAGGCATGTGAGGCTGTGGCGGGGCGGAGG 800
RmRV-trunc1     721    ..........................................-..................................... 799

CmRV-trunc1     801    GTTTCGAGCTCATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGC 880
RmRV-trunc1     800    .....C.......................................................................... 879

CmRV-trunc1     881    ACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATGTGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGT 960
RmRV-trunc1     880    ................................................................................ 959

CmRV-trunc1     961    TCGCGCTTTCCCCTGATTTGTGTCGTTTTAAAAATAGACCTAATGTGTTGCAAGAGCAAATACTAGGTGTTTAACGTGAG 1040
RmRV-trunc1     960    .................C.............................................................. 1039

CmRV-trunc1     1041   TTAGAACACAGATCTTTTTCTCCGGTTTATAACTTATGCTTAAACTAAAGATTTTATCAGGTCGTGGTTCTCTATCTTGT 1120
RmRV-trunc1     1040   .......................C..................................................G..... 1119

CmRV-trunc1     1121   AATCCCAGCGGTCTGCGAGATATGGCACGCGATGTTTATGTTAATTATATGGTGAACTTCTCTTTACTAATCTCTTAGCT 1200
RmRV-trunc1     1120   ................................................................................ 1199

CmRV-trunc1     1201   TTTGTTGTTGTTGTTGTTCTATAAAAATTGTTTTTCTTTTCCCCCTGGGAGCCAATTGTGGACATTTATTAACAGCAAAC 1280
RmRV-trunc1     1200   ...T............................................................................ 1279

CmRV-trunc1     1281   AGCCCGCTGGAGTGCATGTAGGGATCTCGGCTTACTGCAGCCTCTGCCTCCTGCCTTCAAGCTATTCTCCTGCCTCAGCC 1360
RmRV-trunc1     1280   ................................................................................ 1359

CmRV-trunc1     1361   AGAGACTACTCAGCCTCCGGAGTAACTGGGATTGCAGGCGCAGGCCACCAAAAAATTAGGCCCGGCTAATTTTTTGAATT 1440
RmRV-trunc1     1360   .............................................T.................................. 1439

CmRV-trunc1     1441   TTTGGTAGAGACAGGGTTTCGCCGGGCCAGGCGCAGTGGCTCACGCCTGTAATCCAAATACTTTGGGAGGCCGAGGAGGG 1520
RmRV-trunc1     1440   ................................................................................ 1519

CmRV-trunc1     1521   TGGACCATGAGTTCAGGATATCGAGACCATCCTGGCTA--ACAGTGAGACCCCATCTCTACCAAAAATACAAAATAAATA 1598
RmRV-trunc1     1520   ......G...............................AC........................................ 1599

CmRV-trunc1     1599   TATGTGTGTGTATATAGATATATATAGATATATATATATAAACCAGGCGTGCCTGTAATCCCAGCTACTCGGGAGACTGA 1678
RmRV-trunc1     1600   ..................................C............................................. 1679

CmRV-trunc1     1679   GGCAAGAAAACCGCTTGAACCCGGGAGGCAGAGGTTGCAGTAAGCTGAGATCGCGCCACTGCACTCCAGCCTGGGTGACA 1758
RmRV-trunc1     1680   ................................................................................ 1759

CmRV-trunc1     1759   GAGCGAGACTCCGTCTC--AAAAAAAAAAAAAAAAAGAGACAGGGTTTCGCCATGTTGGTCAGGCTTATCTTGAACTCCT 1836
RmRV-trunc1     1760   .................AA..............................A.............................. 1839

CmRV-trunc1     1837   GACCTCAAGTTATCCTCCCGCCTGGGCCTCCCAAATGTTGGGATTACAGGCCTGAGCCACCGCGCCCGGCCTGTTTACAG 1916
RmRV-trunc1     1840   ................................................................................ 1919

CmRV-trunc1     1917   GCTTTACAGACACATTTTGTTTACAAACTTTACATGCACAATTCAATTGCAGTTAGTGTTCATTTCTTTTTTTTTTTTCG 1996
RmRV-trunc1     1920   ........................................G.........................-............. 1998

CmRV-trunc1     1997   TTTTTTTGTTTTTGTTTTTGTTTTTCCCAGTGTGGCCCAAGGAAGCCAAAATGTAGACAACCCTGCTTTAGAGAGTTTGA 2076
RmRV-trunc1     1999   ................................................................................ 2078

CmRV-trunc1     2077   TTCATGGTTCCACTGGGTTATACTTATTGCCTATATTTTGCCTATTATTTGATATTTTCCAAATTTGATTTTTTATTTAT 2156
RmRV-trunc1     2079   ..............................................................................-. 2157

CmRV-trunc1     2157   TTTTTATTTATTTTTATTTTTATTTTTTTTGAGACGGAGTCTCGCTCTGCCACCCAGGCTGGAGTGCAGTGGCCGGATCT 2236
RmRV-trunc1     2158   ...A.T.......................................................................... 2237

CmRV-trunc1     2237   CAGCTCACTGCAAGCTCCGCCTCCCAGGTTCACGCCATTCTCCTGCCTCAGCCTCCCGACTAGCTGGGACTACAGGCGCC 2316
RmRV-trunc1     2238   ................................................................................ 2317

CmRV-trunc1     2317   CGCCACCTCGCCCGGCTAGTTTTTTGTATTTTTTAGTAGAGACGGGGTTTCACCGTGTCAGCCAGGATGGTCTCGATCTC 2396
RmRV-trunc1     2318   ................................................................................ 2397

CmRV-trunc1     2397   CTGACCTTGTGATCCGCCCGTCTCGGCCTCCCAAAGTGCTGGGATTACAGGCTTGAGCCAACGCGCCCGGCCTCAAATTT 2476
RmRV-trunc1     2398   ............................................................C................... 2477
```

```
CmRV-trunc1    2477  GATTTTTTAAAAAGCGTTTTCCTGAAGCGTCCTCAAGCTTTCCTAATAATATGGGAACTTAAACCAGATTGGATGTAACA 2556
RmRV-trunc1    2478  ........T......A.............C.................................................. 2557

CmRV-trunc1    2557  AAACCCAACTGGAAATCTGCAATTTCTAACCTTTATGTTAGATCAAGAAGCAAAACACCCACCAGTCCAAGGGTCAGGGA 2636
RmRV-trunc1    2558  ................................................................................ 2637

CmRV-trunc1    2637  AGCTGAAGATCATGCCTGAGCTGTAACAACACCCTTCCCAGTATACAAAAGCCTATAACAGCCACTTGAATTCTATCTTC 2716
RmRV-trunc1    2638  ................................................................................ 2717

CmRV-trunc1    2717  CAAATGTCTCCTGTGGCCCAGGCTAACCCAGAGCTATCTAAAGAGGCAAATCTGGGAAATGTGCATTCAGCTTAGCAAAA 2796
RmRV-trunc1    2718  ................................................................................ 2797
                5’ flanking ←→ 5’ LTR
CmRV-trunc1    2797  CTGACACTGAAACTAGGCCTCATAAAACTTAAAAAATTAACACCAGATGGCCATGGATAGCCCCCCAATGTTCCCGGAAC 2876
RmRV-trunc1    2798  ................................................................................ 2877
                   Target site duplication
CmRV-trunc1    2877  TCAAACAGAAAAACAATTCCAGGAATGTACATCAGGGGGTCCAGCCCTAAACAGTTCTAGGAAAGTATGGCATGAGGTCC 2956
RmRV-trunc1    2878  ....................................................................C........... 2957

CmRV-trunc1    2957  AGCCCTAAACCATTCCAGGAAAGAAACCCATAAGCTCAAGCCCCGACCAATCAGAACCAGACACCTTGCTCAGGCCATAG 3036
RmRV-trunc1    2958  ...............................................................T................ 3037

CmRV-trunc1    3037  CCAGACCCAATCATCTTGCGCCTTAAGCTTTGTTTGAACTTCGCGCCAAAAGCTGTGTTTGAACTTGTGTTTGCCTATAT 3116
RmRV-trunc1    3038  ................................................................................ 3117

CmRV-trunc1    3117  AAACAGCCTGTAATAAGCGTTCGGGGTCCCAGGGGCCAACTTAGAGCTTGGGACCCTAGTGCGCTAGCAATAAATAACTC 3196
RmRV-trunc1    3118  ................................................................................ 3197
                                                                                        5’ LTR ←┐
CmRV-trunc1    3197  TCTGCTGTGAATCTCGTGTCGGTGATCCTTCACGGCGACCCCTGCCCAGGAGGGAATCGAGAGTTCGGTTCCAACATTTG 3276
RmRV-trunc1    3198  ................................................................................ 3277

CmRV-trunc1    3277  GTACGTTGGCCGGGAAGTGGGGTCGTCCGAGGACCCCCGACCCATCCGGCGGAGACCTATCTGGCCCGGGCCATGGACTG 3356
RmRV-trunc1    3278  ..G.......T.................A................................................... 3357

CmRV-trunc1    3357  CTGACTGAACGGACCTACCAGGTACTTTCGTTTTGTTCTGTCTGTCTTGCCGGCTAACTCTGAACTCTGGGGAGTACTCC 3436
RmRV-trunc1    3358  ..........A............................A...........A............................ 3437

CmRV-trunc1    3437  TTCTGAATTAAGTGGGGAAGGGGGACAGATGTGTCCGGCACCTTCCCACTTTCCCCCCGGGGGACGCCCTGGCGGTAGTC 3516
RmRV-trunc1    3438  ...............................................C................................ 3517

CmRV-trunc1    3517  TGGGAGAAGGCTGACGACTCAGTCAGCCTCCTTAAATCTGTAGGCAGGTTGCCTCTGCCGTCTGAGTATTTTGTGATCTT 3596
RmRV-trunc1    3518  .....................................................C.......................... 3597

CmRV-trunc1    3597  GTGGCGCCACTCTCTGGCCGCGCGGCTTCTCCTTACTTGTCTGGTCTTTGTTTTTCTTACTTTCGTTAAGTCCTTGTCAT 3676
RmRV-trunc1    3598  .......................................................G.....................T.. 3677
                                ┌→ gag-pol
CmRV-trunc1    3677  ACGTGGACGAAATGGGACAGACGTTGACGACTCCTTTGTCTCTGACCCTGACTCACTTCCCTGACGTCCGGGCTCGAGCC 3756
RmRV-trunc1    3678  ................................................................................ 3757

CmRV-trunc1    3757  CACAACCTCTCCGTAGAAGTCCGTAAGGGACGATGGCAAAAATTCTGCTCGTCCGAGTGGCCTACCCTCCATGTTGGGTG 3836
RmRV-trunc1    3758  ................................................................................ 3837

CmRV-trunc1    3837  GCCCCGGGACGGAACATTTGACCTCCCAATTATCTTACATGTTAAAGCAACAGTGATGGATCCTGGGCCACACGGACACC 3916
RmRV-trunc1    3838  .......................................G........................................ 3917

CmRV-trunc1    3917  CAGACCAGGTGGCCTACATAATTACTTGGGAGGATCTGGTCCGAAACCCTCCCTCTTGGGTGAAACCCTTCCTCCATTCC 3996
RmRV-trunc1    3918  ................................................................................ 3997

CmRV-trunc1    3997  CCTTCCCCACCCCAGTCTACCCTTCTTGCCTTAGAAGCCCCGAAGAATCGGAAACCGGACCCGCATAAGCCAGTCCTCCC 4076
RmRV-trunc1    3998  .....................................................T.......................... 4077

CmRV-trunc1    4077  AGATGAACCCCAGAGGGATCTCCTCCTTCTTGACCCCCTGCCTCCTCCACCTCAAAACCCCCTTCTGAGACCTCCACCTT 4156
RmRV-trunc1    4078  ................................................................................ 4157

CmRV-trunc1    4157  ACGCTTCACCCTTGCCCCCTGCCTTGTCCCCAGCCCTTTCCTCTACCGCCTCGGCCCCTACCCTTTCTCCAACTTCCCCC 4236
RmRV-trunc1    4158  ................................................................................ 4237

CmRV-trunc1    4237  TCGGCCCCTCCCTCGACCCCGTCTCCTTCTCCAGCCCCGCCCGAACTCACCCCTCGGACGCTGCCACCGACACCTCGTCT 4316
RmRV-trunc1    4238  ....................A........................................................... 4317

CmRV-trunc1    4317  CCGCTTGCGGCGGACTGAGGACCCAGATGGCCCTTCCACTTGGCAATCCTCCCTTTTTCCCCTCCGTACCGTCAATCACA 4396
RmRV-trunc1    4318  ................................................................T............G.. 4397

CmRV-trunc1    4397  CGGTCCAGTACTGGCCCTTCTCTGCCTCTGACCTCTACAACTGGAAAACTCATAACCCTTCCTTTTCCCAAGACCCTCAG 4476
RmRV-trunc1    4398  ................................................................................ 4477

CmRV-trunc1    4477  GCCCTAACCTCGTTGATAGAATCCATTCTCCTCACCCACCAGCCCACCTGAGATGATTGCCAGCAACTCTTGCAGGTCCT 4556
RmRV-trunc1    4478  ................................................................................ 4557

CmRV-trunc1    4557  CCTAACC--ACTGAAGAAAGGCAGCGAGTCCTCCTGGAGGCCCGGAAAAATGTGCCAGGGCCAGGAGGCCTCCCAACCCA 4634
RmRV-trunc1    4558  .......TA....................................................................... 4637

CmRV-trunc1    4635  ACTTCCCAACGAAATAGACGAGGGATTTCCCCTCACCCGCCCGGACTGGGACTATGAAACGGCACCAGGTAGGGAGAGTC 4714
RmRV-trunc1    4638  ................................................................................ 4717

CmRV-trunc1    4715  TCCGAATCTATCGCCAGGCTCTGTTGGCGGGTCTCAAAGGGGCAGGAAAACGCCCCACAAATTTGGCCAAGGTAAGAACC 4794
RmRV-trunc1    4718  ................................................................................ 4797

CmRV-trunc1    4795  ATAATTCAGGAAAGGGACGAAAGCCCGGCAGCCTTTATGGAGAGGCTTCTGGAAGAGTTCCGAATGTATAGCCCATTCGA 4874
RmRV-trunc1    4798  .......................................................G........................ 4877

CmRV-trunc1    4875  TCCAGAGGCCCTAGAACATAAGGCTACCGTGGTCATGGCGTTCATAGACCAAGCTGCATCCGATATCAAAGGAAAACTCC 4954
RmRV-trunc1    4878  ................................................................................ 4957
```

```
CmRV-trunc1    4955  AAAGGTTAGATGGGATCCAGACTTATGGATTACAGGAATTGGTTAGGGAGGCAGAAAAAGTATATAATAAGAGAGAAACC 5034
RmRV-trunc1    4958  ................................................................................ 5037

CmRV-trunc1    5035  CCTGAGGAAAGGGAAGCCAGGTTAGCGAAAGAACAAGAGGAGCGAGAGGATCGAAGGGATCGAAAGAGGGATAAGCATTT 5114
RmRV-trunc1    5038  ................................................................................ 5117

CmRV-trunc1    5115  AACAAAAATCCTGGCGGCAGTAGTAACAGAGAAAAGACCAGGGAGAGAGGGAGAGAAGCGGAGGCGGCCAAAAGTGAAAA 5194
RmRV-trunc1    5118  ...............A................................................................ 5197

CmRV-trunc1    5195  AAGACCAGTGTGCCTACGCAAAGAACGGGGACGTTAGATTAAAGATTGCCCCAAGCGTCCTAAAGACCAGAAGAAACCTG 5274
RmRV-trunc1    5198  ................................A............................................... 5277

CmRV-trunc1    5275  TCGCTGTCCTCACCCTAGGTGAAGATAGCGAATAGGGGTGTCAAGGCTCTGGAGCCCCCCCCGAGCCCCGGCTAACTCTC 5354
RmRV-trunc1    5278  ................................................................................ 5357

CmRV-trunc1    5355  TCTGTAGGGGGGCACCCTACCACCTTCTTGGTGGATACAGGGGCCCAACATTCAGTTTTGACAAAGGCAGACGGGCCCCT 5434
RmRV-trunc1    5358  ................................................................................ 5437

CmRV-trunc1    5435  GTCTTCTCGCATATCCTGGGTACAGGGGGCAACAGGGGGAAAATTGCACAAGTGGACTAACCACCGGACAGTTAATCTTG 5514
RmRV-trunc1    5438  .................................................G...............A.............. 5517

CmRV-trunc1    5515  GACAAGGAATGGTAACACATTCCTTCTTGGTGGTACCTGAATGCCCATACCCCCTCCTAGGGCGAGATCTTCTGACCAAG 5594
RmRV-trunc1    5518  ................................................................................ 5597

CmRV-trunc1    5595  CTCGAAGCCCAAATCCACTTCTCCGAGACAGGGGCCCAGGTATTAGATCGGGACGGTCAGCCCATCCAAATCTTAACTGT 5674
RmRV-trunc1    5598  ........................A....................................................... 5677

CmRV-trunc1    5675  GTCTCTGCAAGATGAGTATCGGCTTTTTGACGCTCCGGTCATCACTAGCCTCCCTGATGTTTGGTTGCAAGACTTTCCCC 5754
RmRV-trunc1    5678  ................................T............................................... 5757

CmRV-trunc1    5755  AAGCTTGGGCGGAAACGGGAGGACTTGGGCTGGCCAAGTATCAAGCCCCAATCATAATTGATCTAAAGCCCACGGCGGTG 5834
RmRV-trunc1    5758  ................................................................................ 5837

CmRV-trunc1    5835  CCCGTGTCTATTAAGCAATATCCCATGAGCCGAGAGGCTCATATAGGAATTCGGCAGCACATTAACAAATTTCTAGAACT 5914
RmRV-trunc1    5838  ................................................................................ 5917

CmRV-trunc1    5915  CGGAGTGTTGCGACCTTGTCGCTTGCCCTGGAACACTCCTCTTCTGCCAGTAAAAAAGCCTGGTACTCAGGATTACAGGC 5994
RmRV-trunc1    5918  ................................................................................ 5997

CmRV-trunc1    5995  CTGTCCAAGACTTGAGAGAAATTAACAAAAGGACCATGGATATCCATCCCACGGTCCCCAACCCTTACAATTTGCTCAGC 6074
RmRV-trunc1    5998  ................................................................................ 6077

CmRV-trunc1    6075  ACCTTGAGACCAGACTACAACTGGTATACAGTGTTAGATTTAAAAGATGCTTTCTTCTGTTTACCTCTGGCCCCCCAAAG 6154
RmRV-trunc1    6078  ................................................................................ 6157

CmRV-trunc1    6155  CCAAGAACTCTTTGCCTTTGAGTGGAAGGACCCTGAGAAAGGAATTTCGGGCCAATTGACCTAGACCCGGCTTCCCCAAG 6234
RmRV-trunc1    6158  ..........................................................G..................... 6237

CmRV-trunc1    6235  GATTCAAGAACTCTCCCACTCTCTTCGATGAGGCTCTTCATCGAGACCTGACCGACTTCCGGACCCAAAATCCAGAAGTG 6314
RmRV-trunc1    6238  ................................................................................ 6317

CmRV-trunc1    6315  ACCCTACTCCAGTATGTAGATGACCTCCTCCTGGCTGCCCCTACAAAAGAAACCTGTATACAAGGTACCAGGCATCTACT 6394
RmRV-trunc1    6318  ................................................................................ 6397

CmRV-trunc1    6395  CCAGGCACTGGGTGAAAAAGGATACCGGGCATCCGCCAAGAAGGCACAGCTCTGTCAGACCAAGGTAACATACCTGGGGT 6474
RmRV-trunc1    6398  ......................................T......................................... 6477

CmRV-trunc1    6475  ATATCCTGAGTGAAGGGAAAAGGTGGCTCACCCCTGGGCGCATAGAGACAGTGGTTCGCATTCCACCACCACGAACGCCC 6554
RmRV-trunc1    6478  ................................................................................ 6557

CmRV-trunc1    6555  AGAGAGGTATGTGAATTCCTGGGGACTGCTGGGTTCTGTCGCTTGTGGATACCTGGTTATGCTGAACTGGCCGCCCCCCT 6634
RmRV-trunc1    6558  .........C...................................................................... 6637

CmRV-trunc1    6635  TTATGCACTCACCAAGGAAAGTACCCCTTTCACCTGGCAGACAGAGCATCAATTGGCTTTTGAGGCACTAAAAAAGGCAC 6714
RmRV-trunc1    6638  ................................................................................ 6717

CmRV-trunc1    6715  TCTTGTCTGCCCCAGCCCTTGGGTTACCGGACACCTCAAAGCCCTTTACCCTCTTCCTGGAGGAGAGGCAAGGGATTGCC 6794
RmRV-trunc1    6718  .................................T...........................C.................. 6797

CmRV-trunc1    6795  AGAGGAGTCCTGACCCAAAAATTGGGACCTTGGAAAAGACCGGTAGCATACCTGTCTAAAAAGCTGGACCCTGTGGCGGC 6874
RmRV-trunc1    6798  .A.............................................................................. 6877

CmRV-trunc1    6875  CGGCTGGCCCCCGTGTCTTCGTATCATGGCAGCCACCGCTATGCTGGTCAAGGACTCTGCTAAATTGACCCTTGGGCAGC 6954
RmRV-trunc1    6878  .T.......................G...................................................... 6957

CmRV-trunc1    6955  CACTAACTGTTATTACCCCACATGCTCTAGAGGCCATAGTGCGGCAGCCCCCGGACCGGTGGATAACCAACGCACGCCTA 7034
RmRV-trunc1    6958  ............................................T................................... 7037

CmRV-trunc1    7035  ACCCACTATCAGGCCCTCCTACTGGACACGGATCGCGTCCAGTTTGGCCCTCCGGTCACCTTAAACCCTGCTACGCTGCT 7114
RmRV-trunc1    7038  ................................................................................ 7117

CmRV-trunc1    7115  GCCGGTACCAGAAGACCAACCAAGCCCACACGACTGTCGGCAGATATTGGTGGAGACCCATGGAACGCGGGAAGACCTTA 7194
RmRV-trunc1    7118  ................................................................................ 7197

CmRV-trunc1    7195  AAGACCAAGAACTCCCAGACGCGGATCACACCTGGTACACAGACGGCAGCAGTTACCTTGACTCAGGTACCCGGAGGGCG 7274
RmRV-trunc1    7198  ................................................................................ 7277

CmRV-trunc1    7275  GGAGCGGCAGTAGTAGATGGCCACAACACCATTTGGGCACAATCACTACCTCCTGGCACGTCTACACAGAAGGCTGAGTT 7354
RmRV-trunc1    7278  ................................................................................ 7357

CmRV-trunc1    7355  AATAGCACTAACCAAGGCCCTAGAGCTGTCCAAGGGAAAGAAAGCTAACATTTATATTGATAGCCGATATGCCTTTGCAA 7434
RmRV-trunc1    7358  ................................................................................ 7437
```

```
CmRV-trunc1    7435  CGGCTCATACTCATAAAAATATTTATTAAAAAAGAGGTCTCCTAACTTCAGAAGGAAAAGAAATCAAGAACAAAGCTGAA 7514
RmRV-trunc1    7438  ................................................................................ 7517

CmRV-trunc1    7515  ATAATTGCCTTATTAAAAGCCCTTTTTCTTCCTCAAGAAGTGGCCATAATTCACTGCCCCGGACACCAGAAAGGACAGGA 7594
RmRV-trunc1    7518  ................................................................................ 7597

CmRV-trunc1    7595  TCCAGTAGCAGTAGGAAACAGACAGGCTGACCAAATGGCAAGGCAGGCCGCCATGGCGGAAGTACTGACCCTAGCCACAG 7674
RmRV-trunc1    7598  ....................................................G........................... 7677

CmRV-trunc1    7675  AACCTGACGAAACCAGCCACATAACCATTGAACATACCTATACCTCAGAAGACCAGGAAGAAGCAAGAGCCATAGGGGCT 7754
RmRV-trunc1    7678  ................................................................................ 7757

CmRV-trunc1    7755  ATAAAAAACAAAGACACTAAAAACTGGAAAAAAGGAGGAAAGATAGTCCTTCCCCAAAAAGAGGCCCTGGCAATGATCCA 7834
RmRV-trunc1    7758  ................................................................................ 7837

CmRV-trunc1    7835  GCAGATGCATGCCTGGACACACTTGAGTAATCGAAAGCTGAAATTGTTGATTGAAAAAACTGACTTTCTAATCCCAAGGG 7914
RmRV-trunc1    7838  ................................................................................ 7917

CmRV-trunc1    7915  CAAGTACTCTCGTAGAACAAGTAACATCTGCCTGTAAGGTCTGCCAGCAGGTAAACGCTGGGGCTACCCGAGTACCAGCA 7994
RmRV-trunc1    7918  ................................................................................ 7997

CmRV-trunc1    7995  GGGAAACGAGCTCGTGGTAACCGCCCAGGGGCCTATTGGGAAATAGACTTCACTGAAGTAAAACCTCACTATGCTGGATA 8074
RmRV-trunc1    7998  ................................................................................ 8077

CmRV-trunc1    8075  TAAGTACTTACTGGTGTTTGTAGATACCTTTTCAGGATGGGTAGAAGCCTACCCCACCCGGCAAGAAACGGCACACATAG 8154
RmRV-trunc1    8078  ................................................................................ 8157

CmRV-trunc1    8155  TAGCCAAGAAAATTCTGGAAGAAATCTTTCCTAGATTCGGACTCCCCAAGGTAATTGGGTCAGATAACGGGCCGGCCTTT 8234
RmRV-trunc1    8158  ...............................................................................C 8237

CmRV-trunc1    8235  GTTTCTCAGGTAAGTCAGGGGCTAGCCAGGATATTGGGGATTAATTGGAAATTGCATTGTGCCTATAGACCCCAGAGCTC 8314
RmRV-trunc1    8238  ............................................................................A... 8317

CmRV-trunc1    8315  AGGACAGGTAGAAAGGATGAATAGAACAATAAAAGAGACCCTTACTAAATTGACCTTAGAGACTGGTTTAAAAGATTGGA 8394
RmRV-trunc1    8318  ................................................................................ 8397

CmRV-trunc1    8395  GACGCCTCCTATCCCTAGCTTTGTTAAGAGCCCAAAATACGCCTAACCGCTTTGGGCTCACTCCATATGAAATCCTCTAC 8474
RmRV-trunc1    8398  .......................................A.....................C.................. 8477

CmRV-trunc1    8475  GGAGGACCTCCCCCTTTGTCAACCTTACTTGACTCCTTCTCCCCCTCCGATCCTAAGACTGATCTACAGGCCCGGCTAAA 8554
RmRV-trunc1    8478  ................................................................................ 8557

CmRV-trunc1    8555  AGGACTCCAAGCAGTACAGGCCCAAATCTGGGCCCCCTTGGCAGAACTGTACCGACCAGGACATCCACAGACCAGCCACC 8634
RmRV-trunc1    8558  ..........................................................A........T............ 8637

CmRV-trunc1    8635  CCTTCCAGGTGGGGGACTCTGTCTACGTTAGGCGGCACCGCTCTCAAGGACTAGAGCCTCGGTGGAAAGGACCCTACATT 8714
RmRV-trunc1    8638  ................................................................................ 8717

CmRV-trunc1    8715  GTTCTCCTGACCACACCCACAGCCATAAAGGTTGACGGAATCTCCACTTGGATCCACGCATCCCACGCCAAGGCTGCTCC 8794
RmRV-trunc1    8718  ................................................................................ 8797
                                                   → env
CmRV-trunc1    8795  AGAGACGCCCGGACCAACACCACCTGAGACATGGAAACTCCGACGCTCCAAGGACCCGCTCAAGATAAGACTCTCTCGTA 8874
RmRV-trunc1    8798  ................................................................................ 8877
                gag-pol ←
CmRV-trunc1    8875  TCTAACCCCTTGCTTATTGCTTGCCCTCCTTCCCTGCGCCGCTGGCAGTAATAACCCACGCGGACATATGAAACCGCCCG 8954
RmRV-trunc1    8878  ................................................................................ 8957

CmRV-trunc1    8955  GATATATATTCATTGATAATAACCCCCACCGGCCATATAACCTGACCTGGCAGGTAATTAATTTTAATAATCATGAGGTC 9034
RmRV-trunc1    8958  ................................................................................ 9037

CmRV-trunc1    9035  CTAGGTGAAACCTCAAAAATTGCTCCTATAGGGACTTGGTTCCCTGACCTCTATTTCAACCTAGATAAAGTAGCAGGGGT 9114
RmRV-trunc1    9038  ................................................................................ 9117

CmRV-trunc1    9115  AAATGAAATGGAGGGGGGAGAATGGAGAAAACAGGCTAGAAGAGTGTCCATCAGCCGGAACGGGTTCTATGCCTGTCCCG 9194
RmRV-trunc1    9118  ................................................................................ 9197

CmRV-trunc1    9195  GATTTAGGACAGGGGATATGGAAAAAACTTGTGGAGATATAACTCACCTGTATTGCTATAGTTGGTCCTGTGTAACTAAT 9274
RmRV-trunc1    9198  ................................................................................ 9277

CmRV-trunc1    9275  AATGATGGGGAATGGAAATGGGCCACGAAACCCTGGTACATAACCATGTCCTTTGTCCAGCCCTGCACCAGAACCCGATA 9354
RmRV-trunc1    9278  ................................................................................ 9357

CmRV-trunc1    9355  TTCAAAAAATTGCAATCTGGTCCGCATCAAGTTTGAGGATGCGGCAAAATCTGATAATAGTTGGATAACAGGGTTAATAT 9434
RmRV-trunc1    9358  ................................................................................ 9437

CmRV-trunc1    9435  GGGGCCTTTATCTATACCAGAAGCCACTGTATGGGATCCCTATTCAAATTAGATTACTAGTTGACCCGGACATAGCCCCT 9514
RmRV-trunc1    9438  .............................................................C.................. 9517

CmRV-trunc1    9515  GTTGCAGTAGGGCTGAACCAGGTCTTATCAGAAGAAAAGAAGCCCCCGGTTCCCATACCGGAGAAGCCCCAACCAAAAGC 9594
RmRV-trunc1    9518  .............C.................................................................. 9597

CmRV-trunc1    9595  TCCTCAAAGCACTTCCCCGCCCTTGATCTCCGCCTCCTCTGAATACACATCTTCAGCCCAAAACGTCACCCGCTAGCCCC 9674
RmRV-trunc1    9598  ...............................A................................................ 9677

CmRV-trunc1    9675  TTGACCCGGGAATAGGAGACAGGCTCCTAAATCTCATAAAGGGCTCCTATCTTGCCTTAAACCAGACAAGGCCAGAATTC 9754
RmRV-trunc1    9678  ................................................................................ 9757

CmRV-trunc1    9755  ACCTCCTCTTGCTGGCTATGTCTAGCAACAGGCCCTCCTTATTACGAAGGCATCGCCTCCACTAATAAATTCATTGATTC 9834
RmRV-trunc1    9758  ..............C................................................................. 9837

CmRV-trunc1    9835  CACCAATCCTACTGGATGTGCGTGGGAACAACACAATAAACTGACCCTGGCTGAAGTTTCTGGGGTGGGAACCTGTATAG 9914
RmRV-trunc1    9838  ..AT............................................................................ 9917
```

```
CmRV-trunc1    9915  GCCAAGTGCCCCCTAGTCATCGGCATCTTTGTAATATAACCTTGGCAGTACCCAGCTCCAATCGCTATTTGGTCCCCTCT 9994
RmRV-trunc1    9918  ...................C.....................C..................G................... 9997

CmRV-trunc1    9995  GAGACGGACTGGTGGGCTTGCGACACTGGGCTCACCCCCTGTATATCCACAACTGTTTTTAGCAGCGGCACCCACTATTG 10074
RmRV-trunc1    9998  ..................................T...............................A............. 10077

CmRV-trunc1    10075 CGTGTTGGTACAAGTTGTTCCCCGAGTCTACTATCACTCTGGAGACTCCTTGGACCTCCGGTATGAGCAGAAAACTCATA 10154
RmRV-trunc1    10078 ................................................................................ 10157

CmRV-trunc1    10155 TTAGACCTAAAAGAGAACCTATCTCCCTCACCCTCGCCGTCATGCTAGGAATTGGGGTAGCGGCTGGAGTTGGGACCGGG 10234
RmRV-trunc1    10158 ................................................................................ 10237

CmRV-trunc1    10235 ACGGCAGCCCTAGTACATGGTAACTACCATCTGCACCAACTTAGAATAGCCATAGATGAGGACCTTAGAGCCATAAAACA 10314
RmRV-trunc1    10238 ................................................................................ 10317

CmRV-trunc1    10315 CTCTATCACAAAACTTGAAGAGTCTTTGACTTCTCTGTCTGAAGTTGTGTTACAAAACCGACGAGGACTAGAGATTGTCT 10394
RmRV-trunc1    10318 ................................................................................ 10397

CmRV-trunc1    10395 TTCTAAAAGAAGGCGGGCTCTGTGCAGCCCTCAAAGAGCAATGTTGTTTTTATGCAGATCATTCAGGAGTAGTTAAAGAT 10474
RmRV-trunc1    10398 ................................................................................ 10477

CmRV-trunc1    10475 TCTATGGCAAAACTGAGAGAAAGATTAGACAAAAAAAAAAAGGAGAGAAAATCTCAACAAAGTTGGTTTGAAAATTGGTA 10554
RmRV-trunc1    10478 ................................................................................ 10557

CmRV-trunc1    10555 CAACCAATCCCCTTGGCTTAGCACCCTAATCTCCACCATCTTAGGACCCCTCATCCTGCTCACGCTCATCTTGACTTTCA 10634
RmRV-trunc1    10558 ..............................................................T................G 10637

CmRV-trunc1    10635 GGCCGTGCATACTCAACCGCTTACTCACCCTTATTAAAAATAGATTAAATATAGTACATGCTATGGTTCTGACCCAACAA 10714
RmRV-trunc1    10638 ................................................................................ 10717
                                                     env ←┐                           ┌→ 3’LTR
CmRV-trunc1    10715 TACCAGACCCTCAGGACTGAAGAAGAGGCTCAAGATTGAGCCTCTGACACAAAAAGAGGAGGGAATGAAACTAGGCCTCA 10794
RmRV-trunc1    10718 ................................................................................ 10797

CmRV-trunc1    10795 TAAAACTTAAAAAATTAACACCAGATGGCCATGGATAGCCCCCCAATGTTCCCGGAACTCAAACAGAAAAACAATTCCAG 10874
RmRV-trunc1    10798 ................................................................................ 10877

CmRV-trunc1    10875 GAATGTACATCAGGGGGTCCAGCCCTAAACAGTTCTAGGAAAGTATGGCATGAGGTCCAGCCCTAAACCATTCCAGGAAA 10954
RmRV-trunc1    10878 ..............................................C................................. 10957

CmRV-trunc1    10955 GAAACCCATAAGCTCAAGCCCCGACCAATCAGAACCAGACACCTTGCTCAGGCCATAGCCAGACCCAATCATCTTGCGCC 11034
RmRV-trunc1    10958 ................................................................................ 11037

CmRV-trunc1    11035 TTAAGCTTTGTTTGAACTTCGCGCCAAAAGCTGTGTTTGAACTTGTGTTTGCCTATATAAACAGCCTGTAATAAGCGTTC 11114
RmRV-trunc1    11038 ................................................................................ 11117

CmRV-trunc1    11115 GGGGTCCCAGGGGCCAACTTAGAGCTTGGGACCCTAGTGCGCTAGCAATAAATAACTCTCTGCTGTGAATCTCGTGTCGG 11194
RmRV-trunc1    11118 ................................................................................ 11197
                                                                   3’ LTR ←┬→ 3’ flanking
CmRV-trunc1    11195 TGATCCTTCACGGCGACCCCTGCCCAGGAGGGAATCGAGAGTTCGGTTCCAACAACACATTACAAATCCACCACAGTTCA 11274
RmRV-trunc1    11198 ................................................................................ 11277
                                                                      Target site duplication
CmRV-trunc1    11275 CTTCTCAACATGGCATCAATGTACACCTTCCTCCTCCTCTTCTTTTCCTCCTCTTCCTCCTTTCTCTTCTTCCTCCTCCT 11354
RmRV-trunc1    11278 ............................---................................................. 11354

CmRV-trunc1    11355 TTTTCCTCCTTGTAAAACCATACTTAAGGTTCAAATAAAGACAATAGGAAAATCATACTTTTTCTCATTCAGTGAAAAGA 11434
RmRV-trunc1    11355 ..............................................................................T. 11434

CmRV-trunc1    11435 TGCCAGCCCTTTCCCCAACAGGATAAAGTCCTTTATTTGTCTTTGAATGATGTTCACTCCTTTTTTAGCTAGATCACATT 11514
RmRV-trunc1    11435 .....A.......................................................................... 11514

CmRV-trunc1    11515 CCATATTTGCTATCCTGTAACCTAAATATGGAGATTTCTAAAGGTTAATTATTAATGAATATGCTAGCTAGGGTGATGGA 11594
RmRV-trunc1    11515 ................................................................................ 11594

CmRV-trunc1    11595 AGAGGAGAATTCTTAAGAAAAGTATTGGTTAATATGTACCCAAATGTATTCATCTCAAAATAATGAAGAAATACTCATAA 11674
RmRV-trunc1    11595 ................................................................................ 11674

CmRV-trunc1    11675 CTATTACAAGTCCTCATTTCTGCAACTGTCGCTGGCCCTAGCTGGTATTTGTAACTATCTTCACATCTTGCCCTCACAGC 11754
RmRV-trunc1    11675 ........................................................G....................... 11754

CmRV-trunc1    11755 CATTTCCTTTGGGCCTGGGCCGGTGGATGTGAGCGTGGGAAGGATAATAACCCGCAGTGCCCCGCACCGCACCATCCTCT 11834
RmRV-trunc1    11755 ...........................C........................................A........... 11834

CmRV-trunc1    11835 GGTCTCCAGAGGTGGGATCCACAGAGCAGGGACCTTTGCTGAATTCTAAACCAGCTCTGGGAATACAAGAACAGCTGCCT 11914
RmRV-trunc1    11835 ................................................................................ 11914

CmRV-trunc1    11915 TTGAGAAACAGTTCAGGTCTCCATATTCCAGACTTCCTAGTAGGACTAAACGCTTAATGTTCACACGGGAATTCATGGTT 11994
RmRV-trunc1    11915 ................................................................................ 11994

CmRV-trunc1    11995 CTCCTAAAATCCTGCGGTGGGATGCCTTCCTTCACCGAGCATTTACTGGGCGCCTACTAGGTGCTAGGCTGACTCTGGGC 12074
RmRV-trunc1    11995 ....................................C..................................G........ 12074

CmRV-trunc1    12075 TGTGGGCTGCAGCCTGTGCAAACTCCAGGTCCCTGTGCGGTGGATTCCACCTCTGAAGGCCCGAGGGCTGCACGGTGAGG 12154
RmRV-trunc1    12075 .......................................................................G........ 12154

CmRV-trunc1    12155 ACACTACGGGTTCTGGTCCTTCCCACCGCGCGTCCACTGGCCCAGGCCGGGAGGAAATGACTGTGAGCGCAAGGTGTGAA 12234
RmRV-trunc1    12155 ....C........................................................................... 12234

CmRV-trunc1    12235 GCGGGGCTAAGAATGGGAAAGGGAGTAAGTGACTTCCGCCGGAGAGGCTGCAGTTGTCGCCTGTGGAGGGGCCGGCGACT 12314
RmRV-trunc1    12235 ..........................................................C..................... 12314

CmRV-trunc1    12315 TTTGAGACGGTGAAGAGCAGGCGGGGGAAGCAGTGAGGAGCAGACGGTGAGGAGCAGGCGGCCGTTTGGCCGTTGGAGTC 12394
RmRV-trunc1    12315 ................................--------------........................G......... 12380
```

```
CmRV-trunc1    12395 CCCCAGCTGACTCTGGCCGGGTAGAATGTCCAAACGTGCCCAAAGACTCGACTCTTGGCTTCATTGTCTCCCGCGTTCTG 12474
RmRV-trunc1    12381 ................................................................................ 12460

CmRV-trunc1    12475 GGTTTCTGGAGTGCGCGGGATTGGACAGGCCACAAACATATGAGATTATCTGCAGCCAGAGATTAAAAAGCCGTTTCTGG 12554
RmRV-trunc1    12461 ................................................................................ 12540

CmRV-trunc1    12555 TATCAGCAAATGAGTTTAAGACACAACCCATTAGAGTTTCTGTGGCGCAATCGGTTAGCGTGTTCCGCTGTTAATCATAA 12634
RmRV-trunc1    12541 ................................................................................ 12620

CmRV-trunc1    12635 AGGTTGATGGTTCAAGCCCACCCAGGGACGGTGCCCTTTATTTTT-TTTTTTTTTGTTTTTTGTTTTTTTTTTTTTGAGA 12713
RmRV-trunc1    12621 .............................A.........T.....G...G...........---................ 12697

CmRV-trunc1    12714 CGGAGTCTCGCTCTGTCGCCCAGCCCAGGCTGGAGTGCAGTGGCGCGATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGT 12793
RmRV-trunc1    12698 ................................................................................ 12777

CmRV-trunc1    12794 TCACGCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCCCACAACCGCGCCCGGCTAATTTTTTGTAT 12873
RmRV-trunc1    12778 ................................................................................ 12857

CmRV-trunc1    12874 TTTTAGTAGAGACGGGGTTTCACCGTGGTTTCGATCTCCTGACCTTGTGATCCGCCCGCCTCGGCCTCCCAAAGTGCTGG 12953
RmRV-trunc1    12858 .............................................................G.................. 12937

CmRV-trunc1    12954 GATTACAGGCGTGAGCCACCGCGCCCGGCCGGTGCCCTTTATTTAAGCAAGATTCACGCTTTCTGTTTAAAGACTCTAAT 13033
RmRV-trunc1    12938 ................................................................................ 13017

CmRV-trunc1    13034 CCTGGGAATCCTCCGGACCTCTACTAGCTTTAAATCAGATGCCAAAATATGCTGAGCGAAGGCTCAAGTTCCATGGAAGG 13113
RmRV-trunc1    13018 ................................................................................ 13097

CmRV-trunc1    13114 TCCTAAAACAGGAATATTTCAGCAAAAGTAATTTTTCAATAGAGCCTGAGCCTGGGGCTTCAGAGCTTACCCCATATCCA 13193
RmRV-trunc1    13098 ................................................................................ 13177

CmRV-trunc1    13194 CTGCACCTATTTCATCACTTCCTGAAGCTAGGTTTTTACAATTGCACATTTCTGGAAATC 13253
RmRV-trunc1    13178 ............................................................ 13237
```

Figure S12. Alignment of the nucleotide sequence between CmRV-trunc1 and RmRV-trunc1. The proviral sequences with the flanking sequences, 5’ and 3’ LTRs, *gag*, *pol*, and *env* genes are shown. Flanking 4-bp target site duplication (TSD) sequences are shown in gray. Dots indicate identical nucleotide sequences, and dashes indicate the missing nucleotides. Nucleotide sequences encoding Env are highlighted in green.

```
                         5’ flanking
PNRC2-BoRV      1        GAGCTTGAAGTCAGTAAGTAGCGGCTGCGCTAAGGGCGCGGCCATGTTGGTGGCCGCGCTCCCCATTGTTGGGGGAAGGA 80
PNRC2-ChRV      1        ................................................................................ 80

PNRC2-BoRV      81       GGCGAGCGGAGGCGAGATCCGGGTAACCCGGGCGGGCGCGGGCGGTCGGTCGGTCACGGGCGGTCGCAGCGCCGCGACCC 160
PNRC2-ChRV      81       ..........................................----.................................. 156

PNRC2-BoRV      161      CTGCGGCTCCCGAGAGATGGGGCCACTGGGAAGCCAAGCCCTTGTGGCGTCGGAGGAGGCGTTTTCCGGGTTCGAGTTGT 240
PNRC2-ChRV      157      ................................................................................ 236

PNRC2-BoRV      241      GAATGACTCACAGTCCGGAAATGGCAATTTATTTAGGGTTGTTTATCTGTTTGGAAGGCTCTCTTACCCCTTTGAAAGGA 320
PNRC2-ChRV      237      ................................................................................ 316

PNRC2-BoRV      321      GGGGGCTTAACTATTTGGTAACTTATTTTCTTGGAACATCAAATTAAAACCAAACCTTGAAAGTTGAGCTGTTCCCTCTA 400
PNRC2-ChRV      317      ......C......................................................................... 396

PNRC2-BoRV      401      GGGAGACATTTTTATATATTTAAGATGGGCTTAGACTGTTAGGGCTTGTGAGAACTTATTTTAAGACGGGGAGGAGGCGA 480
PNRC2-ChRV      397      ................................................................................ 476
                                                                          5’ flanking ◄──┬──► 5’ LTR
PNRC2-BoRV      481      GAGTGTTCATAACCAAAGGAAACACTGCGTGATGTCTACATGTTCCAAGAGACTGAAAATGGATGTGATGAGAAACCATA 560
PNRC2-ChRV      477      ................................................................................ 556
                                                                             Target site duplication
PNRC2-BoRV      561      GAGAAAATACAGCCACTAAAACACCAGATGGCCAGGAGTCAGGGTGTGTCAATGGCTATCATAAACACCAGATGCCCAGG 640
PNRC2-ChRV      557      ................................................................................ 636

PNRC2-BoRV      641      GTAAGACAGGGAGCCCCCCCCGCCCCTTGTTCCCGGAACTAAAGCAGAAAAACATTTCCAGGAAATACCTCCTCCCGCCT 720
PNRC2-ChRV      637      .............-.................................................................. 715

PNRC2-BoRV      721      ATAAGCCCCTGACCAGCCCCTGACCAACCAGTGTGAGACAGCTAGCTCAGACCATAATTAGAACCAATCAGTTACTTCCA 800
PNRC2-ChRV      716      ................................................................................ 795

PNRC2-BoRV      801      AACCTCGCGCCCAAAAACTGTTCAAATGTGTAACCCAATCAGCTTATTGTAACCTCCCCCTGTTTGAATTTGTGGTTTTT 880
PNRC2-ChRV      796      ................................................................................ 875

PNRC2-BoRV      881      GCCTTTATAAGCAGTCTGTAACAATCATTCGGGGTCTCCTGGCCTTATGTGCTGGGGACCCTAGCGCGCTAGTAATAAAT 960
PNRC2-ChRV      876      ................................................................................ 955

PNRC2-BoRV      961      AGTGTCTCTTTGCTGTGATCTCTGTGTCGAGTGGTCTCTGGCGGCGGCCCCGTCCCGAGGCAGGAATCTTGAGTAAGGTT 1040
PNRC2-ChRV      956      ................................................................................ 1035
                    5’ LTR ◄─┐
PNRC2-BoRV      1041     CCAACATTTGGGGGCTCGTCCGGGATTGGTCGCCGCCTGACGACCCCCGACCCGCACGGCGAAGACCCACTCGGCCCGGG 1120
PNRC2-ChRV      1036     ................................................................................ 1115

PNRC2-BoRV      1121     CCATCGACTGACGACCAGGCACCCTCAGGTACTTTCGTTTTTGTTTTGTTCGTCTATTTGCGTTGCCGGCTAAACCGGAT 1200
PNRC2-ChRV      1116     .....................................-.......................................... 1194

PNRC2-BoRV      1201     GAGTACTCAGTCTGTATAAGTGTGGAAGGGGGGCAGACGTGCCCAGCACCTTCCCACTTACGCCCCGGGGGATGCCCTGG 1280
PNRC2-ChRV      1195     ................................................................................ 1274

PNRC2-BoRV      1281     CGGTTGTCTGGAGGAAAACTGACGATTCTGTCAGTCTCCTCACCTCTGAAGGCAGGTTCTCCCTGCCATCTGAATCCTTC 1360
PNRC2-ChRV      1275     ................................................................................ 1354

PNRC2-BoRV      1361     GTAGGTCTGTGGCGCCGATCCCTCGTCACGCGGCTTCTCTGTGTGTGTGTAGTCTCCGCTTATATCTTTGTTCTCCTTGT 1440
PNRC2-ChRV      1355     ...........................G.................................................... 1434

PNRC2-BoRV      1441     TCTGTCTATCTGTGCGTTAAAGACTTTAGATACAACTATGGGACAGACACTGACGACTCCTCTATCTTTGACCCTGACTC 1520
PNRC2-ChRV      1435     ................................................................................ 1514

PNRC2-BoRV      1521     ACTTCCCTGACGTACGGGCACGAGCCCACAATCTCTCTGTGGAAATTCGCAAAGGGCGATGGCAAGCCTTCTTTCTGCTC 1600
PNRC2-ChRV      1515     ................................................................................ 1594

PNRC2-BoRV      1601     TTCTGAATGGCCCACCCTCGGCGTTGGGTGGCCCCAAGGCGGGACTTTTGACCTCTCAATCATCTTACAGGTCAAGATAA 1680
PNRC2-ChRV      1595     ............................................................T................... 1674
                              ┌─► gag-pol
PNRC2-BoRV      1681     AAGTGATGGATCCAGGGCCACGCGGCCAGCCTGACCAGGTGGCCTATATCATCACCTGGGAAGACCTGGTCCGGGATCCC 1760
PNRC2-ChRV      1675     ................................................................................ 1754

PNRC2-BoRV      1761     CCCCATTGGGTGAAGCCCTTCCTCCCCTCAGCTCCCCCTTCCCAGTCGACCCTCCTCGCCTTGGAAGCCCCCCGAAACCA 1840
PNRC2-ChRV      1755     ................................................................................ 1834

PNRC2-BoRV      1841     AACCCCGGTCCCCCCGAAACCTGTCCTCCCGGATGAGGGTCAGAGGGACCTCCTCCTCCTAGACCCCCTCCCTCCTCCGC 1920
PNRC2-ChRV      1835     .............T.................................................................. 1914

PNRC2-BoRV      1921     CTCACAATCCCCTCCTTAACCCCCCTCCTTACCCCTCGCCCTCAACCCCGCGCTGTCTCCTGCCCCTTCTACCACCCCTT 2000
PNRC2-ChRV      1915     ................................................................................ 1994

PNRC2-BoRV      2001     CGGCTCCTACTCTTTCTCCAACCTCTTTTTCCACCCCCTCTCTGATCCCATCC---CCTCCGGCCCCGCCCGAACTCACC 2077
PNRC2-ChRV      1995     .....................................................CCT........................ 2074

PNRC2-BoRV      2078     CCTCAGACCCCGCCTCAGACACCCCGCCTCCGCTTACGGTGGGCCAAGGACCCTGGCGACCAGTCCACCTGGCAGTCCTC 2157
PNRC2-ChRV      2075     .......................................C........................................ 2154

PNRC2-BoRV      2158     CCTTTTTCCCCTCCGCACCGTAAATCGCACGGTCCAGTATTGGCCCTTCTCGGCCTCAGAGCTCTACAATTGGAAGACCC 2237
PNRC2-ChRV      2155     ............................................................T................... 2234

PNRC2-BoRV      2238     ATAACCCCCCCTTCTCCCAAGATCCGCAGGCCCTGACCTCTCTGATAGAATCCATTCTCCTCACTCACCAACCCACCTGG 2317
PNRC2-ChRV      2235     ................................................................................ 2314

PNRC2-BoRV      2318     GATGACTGTCAACAGCTCTTACAGGTTCTTCTGACCACAGAAGAGAGGCAACGAGTCCTCCTCGAGGCCCGGAAAAATGT 2397
PNRC2-ChRV      2315     ................................................................................ 2394

PNRC2-BoRV      2398     GCCAGGGCCAGGAGGACTCCTGACCCAACTCCCCAACGAAATAGATGAGGGATTTCCCCTCACCCGCCTGGATTGGGACT 2477
PNRC2-ChRV      2395     ....................C........................................................... 2474
```

```
PNRC2-BoRV     2478  ATGAAACAGCAACAGGTAGGGAGAGTCTCTGAATCTATCGCCAGGCTCTGTTGGCAGGTCTCAAAGGGGCCGGAAAGCGC 2557
PNRC2-ChRV     2475  .............................C.................................................. 2554

PNRC2-BoRV     2558  CCCACCAATTTGGCTAAGGTAAGAACTATTACTCAGGAAAAGAACGAAAGCCCGGCAGCCTTCATGGAAAGGCTCCTAGA 2637
PNRC2-ChRV     2555  ................................................................................ 2634

PNRC2-BoRV     2638  GGGGTTTCGAATGTACACTCCATTCGATCCCGAGGCCCCGGAGCACAAGGCCACCGTAGCTATGTCATTCATAGACCAGG 2717
PNRC2-ChRV     2635  ................................................................................ 2714

PNRC2-BoRV     2718  CAGCGCTAGACATAAAGGGAAAGCTCCAAAGATTGGACGGGATCCAGACCTATGGGCTGCAAGAACTAGTTAGAGAGGCA 2797
PNRC2-ChRV     2715  ................................................................................ 2794

PNRC2-BoRV     2798  GAAAGAGTTTATAACAAGAGAGAGACTACTGAGGAAAAAGAAGCTAGGCTAGCAAAGGAACAGGAGGAACGAGAGGATCG 2877
PNRC2-ChRV     2795  ................................................................................ 2874

PNRC2-BoRV     2878  ACGAGATTGTAAGAGGGACAGGCATTTAACTAAAATCCTGGCAGCAGTAGTGACAGGGAAAGGGCCAGGGAGAGAGGGGG 2957
PNRC2-ChRV     2875  .......C........................................................................ 2954

PNRC2-BoRV     2958  GAGAACGAAGGCGCCCAAAAGTGGATAAAGACCAATGTGCCTATTGTAAAGAACGAGGACATTGGATTAAAGATTGTCCT 3037
PNRC2-ChRV     2955  ................................................................................ 3034

PNRC2-BoRV     3038  AAACGCCCTAAGGACCGGAAGAACCCCGGCCCCGTCCTGACCCTGGGAGAAGACAGCGATTAGGGGCGTCAGGGCTCCGA 3117
PNRC2-ChRV     3035  ...............T...................................................A............ 3114

PNRC2-BoRV     3118  AG-CCCCCCCGAGCCCCGGTTAACCCTTTCTATAGGGGGGCGCCCCACCACCTTTCTGGTGGACACCGGGGCCCAGCATT 3196
PNRC2-ChRV     3115  ..C............................................................................. 3194

PNRC2-BoRV     3197  CAGTTTTGACAAAAGCAGGTGGGCCCCTTTCATCCCACACCTCTTGGGTCCAAGGAGCAACAGGAAGGAAGCTACACAAG 3276
PNRC2-ChRV     3195  ................................................................................ 3274

PNRC2-BoRV     3277  TGGACCACCCACCGAACAGTAAACCTTGGAAAAGGTATGGTGACTCATTCTTTCTTAGTAGTACCTGAATGCCCATATCC 3356
PNRC2-ChRV     3275  ................................................................................ 3354

PNRC2-BoRV     3357  CCTTCTGGGGCGGGATCTGTTGACCAAGCTCGGAGCCCAGATACATTTCTCGGAGAGAGGGGCCCAGGTGCTGGGTGAGG 3436
PNRC2-ChRV     3355  ................................................................................ 3434

PNRC2-BoRV     3437  ATGGTCAGCCTATACAAATTCTGACCGTATCCTTGCAAGATGAGTACCGGCTTTTCGAGACTCCTATCTTCACCAGCCCT 3516
PNRC2-ChRV     3435  .............C.................................................................. 3514

PNRC2-BoRV     3517  CCCGATAACTGGCTGCAAGAATTTCCCCAGGCTTGGGCAGAGACAGGGGGACTTGGACTGGCAAAATTTCAAGCCCTGAT 3596
PNRC2-ChRV     3515  ........T....................................T..............................C... 3594

PNRC2-BoRV     3597  TATAGTTGACCTCAAACCCACCGCAGTGCCCGTGTCCATTAAGCAATACCCCATGAGCCGAGAAGCCCGTATGGGCATCC 3676
PNRC2-ChRV     3595  ................................................................................ 3674

PNRC2-BoRV     3677  AACAGCACGTTAATAAATTTCTGGAATTAGGGGTCTTACGGCCATGCCGCTCACCTTGGAATACGCCCCTCCTTCCGGTA 3756
PNRC2-ChRV     3675  .......T.......................................T................................ 3754

PNRC2-BoRV     3757  AAGAAACCTGGGACCCAAGATTATAGGCCCGTCCAGGACTTGAGAGAAATTAATAAGAGAACCATGGACATACATCCCAC 3836
PNRC2-ChRV     3755  ................................................................................ 3834

PNRC2-BoRV     3837  AGTCCCCAACCCTTACAACCTGCTCAGTACCTTGAGACCAGACCACAACTGGTATACAGTACTAGACCTAAAAGATGCAT 3916
PNRC2-ChRV     3835  ................................................................................ 3914

PNRC2-BoRV     3917  TCTTTTGCTTACCCCTGGCTCCCCAAAGCCAAGAGCTTTTTGCCTTTGAATGGAGGGACCCTGAGAGGGGAATCTCAGGC 3996
PNRC2-ChRV     3915  ................................................................................ 3994

PNRC2-BoRV     3997  CAATTAACTTGGACTCGGCTTCCCCAAGGGTTCAAGAATT---CTACCCTCTTTGACGAGGCTCTTCACCGGGACTTGAC 4073
PNRC2-ChRV     3995  ........................................CTC..................................... 4074

PNRC2-BoRV     4074  TGATTTTCGCACCCAGCACCCAGATTTAACTCTGCTCCAGTATGTAGATGACCTCCTCCTGGCCGCCCCCACTAAGGGAG 4153
PNRC2-ChRV     4075  ................................................................................ 4154

PNRC2-BoRV     4154  CCTGCCTACAGGGCACCAGGCAACTGCTCCAGGAGCTCGGAGAAAAAGGATACCGAGCCTCTGCCAAGAAAGCACAAATC 4233
PNRC2-ChRV     4155  ................................................................................ 4234

PNRC2-BoRV     4234  TGCCAGACTAAGGTAACCTACCTGGGATACATCTTAAGTGAAGGAAAAAGGTGGCTCACCCCTGGGCGGATAGAGACTGT 4313
PNRC2-ChRV     4235  ................................................................................ 4314

PNRC2-BoRV     4314  AGCCCGCATTCCGCCACCTCGGAGTCCCAAGGAGGTGCGTGAGTTTCTGGGGCCTGCCGGGTTCTGCCGCCTGTGGATAC 4393
PNRC2-ChRV     4315  ....................................................A........................... 4394

PNRC2-BoRV     4394  CTGGTTTTGCTGAGTTGGCAGCCCCCCTTTATGCCCTCACCAAAGGGAACAACCCCTTTACCTGGCTGGAGGAACACCAA 4473
PNRC2-ChRV     4395  ................................................................................ 4474

PNRC2-BoRV     4474  CAGGCCTTCGAAACTTTAAAGAAGGCACTCCTCTCTGCCCCAGCCCTCGGGCTACCTGACACATCCAAGCCTTTTACCCT 4553
PNRC2-ChRV     4475  ................................................................................ 4554

PNRC2-BoRV     4554  CTATGCAGACGAGAGACGGGGGATAGCCAAAGGGGTCTTAACCCAAAAACTGGGGCCCTGGAAGAGACCGGTAGCCTACT 4633
PNRC2-ChRV     4555  ................................................................................ 4634

PNRC2-BoRV     4634  TGTCTAAGAAATTGGACCCTGTGGCGGCTGGGTGGCCCCCTTGCCTCCGCATTATGGCAGCCACCGCTATGCTAGTCAAA 4713
PNRC2-ChRV     4635  ...........C.................................................................... 4714

PNRC2-BoRV     4714  GACTCTGCTAAGTTAACCCTTGGGCAACCATTGACTGTCATTACCCCGCACGCCTTAGAGGCCATAGTGCGGCAGCCCCC 4793
PNRC2-ChRV     4715  ................................................................................ 4794

PNRC2-BoRV     4794  GGACCGTTGGATCACCAACGCTCGTCTAACCCACTACCAAGCCCTACTACTAGACACGGACCGCGTCCGCTTTGGCCCTC 4873
PNRC2-ChRV     4795  ........................................................T....................... 4874

PNRC2-BoRV     4874  CGGTCACTCTGAATCCTGCCACCTTGCTACCTGTACCGGAGGTCCCGCTAAGCCCCCACGACTGTCGACAAGTGCTGGCG 4953
PNRC2-ChRV     4875  ................................................................................ 4954
```

```
PNRC2-BoRV     4954  GAGACCCACGGGACTAGAGAAGACCTCCAGGACTACGAGCTCCCAGACGCAGACCATACTTGGTACACAGACGGTAGCAG 5033
PNRC2-ChRV     4955  ................................................................................ 5034

PNRC2-BoRV     5034  CTTCATGGACGCAGGTACCCGGAGGGCGGGGGCGGCGGTAGTGGATGGACATGCCACGATATGGGCGCAGGCACTGCCTC 5113
PNRC2-ChRV     5035  ...................T..................A......................................... 5114

PNRC2-BoRV     5114  CCGGAACGTCTGCTCAAAAGGCTGAACTAGTTGCTCTAACAAAGGCCTTAGAGCTGTCGCAGGGGAAAAAGGCTAACATC 5193
PNRC2-ChRV     5115  ................................................................................ 5194

PNRC2-BoRV     5194  TACACAGACAGTCGGTATGCCTTTGCGACAGCCCACACCCATGGGAGCATTTACGAGAGGCGAGGTCTCCTAACATCAGA 5273
PNRC2-ChRV     5195  ................................................................................ 5274

PNRC2-BoRV     5274  AGGAAAAGAAATCAAAAATAAGGCCGAAATAATCGCCTTATTAAAGGCCCTCTTCCTCCCTAAAAAGGTGGCCATAATTC 5353
PNRC2-ChRV     5275  ............A................................................................... 5354

PNRC2-BoRV     5354  ATTGTCCTGGGCATCAAAAAGGACATGACCCCGTCGCCCAGGGTAACAGGCAAGCTGACCAGGCGGCCAAGCAGGCTGCT 5433
PNRC2-ChRV     5355  ................................................................................ 5434

PNRC2-BoRV     5434  AGAATATTGACCTTAGTTTCGGAAACCAAAAAGGCTGACCGAATACCCCCTTCCCCACGTTATACCTACACACCAGAGGA 5513
PNRC2-ChRV     5435  ........C................................G...................................... 5514

PNRC2-BoRV     5514  CCGGGAAGAGGCAATAGCCTTAGGAGCCACAGAAAACCAAGAGACTAAGAATTGGGAAAAAGACGGGAAGACGGTCCTCC 5593
PNRC2-ChRV     5515  .......................................................................T........ 5594

PNRC2-BoRV     5594  CACAAAAACAGGCCACGGCCATGGTGCAGCAGATGCACGCCTGGACACATTTAAGTAGTAAGAAACTAAAACTGCTCATT 5673
PNRC2-ChRV     5595  ................................................................................ 5674

PNRC2-BoRV     5674  GAAAAGACTGACTTCCTAATCCCCAGGGTCGGGACCCTCCTGGAACAAATAACGCTCGCTTGCAAGGCCTGCCAACAAGT 5753
PNRC2-ChRV     5675  ................................................................................ 5754

PNRC2-BoRV     5754  AAACGCCGGGGCCACGCGAGTCCCGGCGGGGATAAGGACACGGGGCAACCGCCCTGGGACCTATTGGGAAGTAGATTTTA 5833
PNRC2-ChRV     5755  ................................................................................ 5834

PNRC2-BoRV     5834  CTGAAATAAAGCCTCACCATGCAGGATATAAATATTTGTTAGTATTTGTAGACACATTTTCAGGATGGGTAGAAGCCTAC 5913
PNRC2-ChRV     5835  ................................................................................ 5914

PNRC2-BoRV     5914  CCCACCCGGCAAGAAACGGCCCACGTAGTGGCCAAGAAGATATTAGAAGAAATTTTCCCCAGGTTCGGACTCCCCAAGGT 5993
PNRC2-ChRV     5915  ................................................................................ 5994

PNRC2-BoRV     5994  AATCGGGTCAGATAATGGGCCAGCCTTCGTCTCCCAGGTAAGTCAGGGACTTGCCAGGATACTGGGGATTGATTGGAAAC 6073
PNRC2-ChRV     5995  ................................................................................ 6074

PNRC2-BoRV     6074  TTCATTGTGCTTATAGGCCCCAGAGTTCAGGACAGGTAGAACGGATGAATAGAACTATTAAAGAGACCTTAACAAAATTG 6153
PNRC2-ChRV     6075  ................................................................................ 6154

PNRC2-BoRV     6154  ACCTTAGAGACTGGCTTAAAAGATTGGAGACGTCTCCTATCCCTAGCTCTCTTGAGAGCCCGAAATACGCCTAATCGCTT 6233
PNRC2-ChRV     6155  ................................................................................ 6234

PNRC2-BoRV     6234  TGGGCTCACCCCTTATGAAATCCTCTATGGGGGACCACCCCCCTTGTCAACCTTGCTTGATTCTTTCGCCCCCTCTAACC 6313
PNRC2-ChRV     6235  ......................................................AA........................ 6314

PNRC2-BoRV     6314  CTAAGACTGACTTGCAGGCTCGACTAGAAGGACTACAGGCGGTGCAAGCCCAAATTTGGGCTCCTTTGGCGGAACTGTAC 6393
PNRC2-ChRV     6315  ................................................................................ 6394

PNRC2-BoRV     6394  CAGCCTGGACACCCACAAACCAGTCATCCTTTCCAAGTGGGAGACTCCGTCTATGTCAGACGACATCGCTCCCAAGGATT 6473
PNRC2-ChRV     6395  ................................................................................ 6474

PNRC2-BoRV     6474  AGAACTCCGGTGGAAGGGACCATACATCGTTCTCCTCACCACACCCACTGCCGTGAAAGTTGACGGGGTCGCCGCCTGGA 6553
PNRC2-ChRV     6475  .....C.......................................................................... 6554

                                                                                         env
PNRC2-BoRV     6554  TCCACGCATCCCACGTAAAAGCCGCTCCGAGGGTGCCAGAATCAGCATCGCCTGAGAAATGGAGACTTCGTCGCTCCAGG 6633
PNRC2-ChRV     6555  ................................................................................ 6634

                                                 gag-pol
PNRC2-BoRV     6634  GACCCCCTCAAGATAAGACTTTCCCGTGTCTAACCCCTCACCTACTGTTAACTCTTTTCCTTCCCTGGGTTGTCGGAAGC 6713
PNRC2-ChRV     6635  ................................................................................ 6714

PNRC2-BoRV     6714  AGCAACCCCCACCAACCCTATCGATTGACTTGGCAAATAACTAATTTTAAAACCCATGAAGTCCTCAACGAGACTTCACA 6793
PNRC2-ChRV     6715  .............................C.................................................. 6794

PNRC2-BoRV     6794  TGTAGCCCCTTTAAATACCTGGTTCCCTGACCTCTACTTCAATCTTGACAAAATAGCCATGACAGATGAAATGGAGGGTG 6873
PNRC2-ChRV     6795  ..............................................................T................. 6874

PNRC2-BoRV     6874  GTGAGTGGAGAAAGCAAGCGAGAAGAGTCTCCCTAAGTCGAAACGGGTTTTATGTCTGCCCTGGATTCCGGACGGGACCG 6953
PNRC2-ChRV     6875  ................................................................................ 6954

PNRC2-BoRV     6954  ATGAAAAAGACCTGTGGTGAAATAATGTCTCTGTACTGTGCAAGTTGGTCATGTGTAACAACTAATGATGGGGAATGGAA 7033
PNRC2-ChRV     6955  .............................C.................................................. 7034

PNRC2-BoRV     7034  ATGGAAAACCCAACCCTGGTATTTAACCATGTCCTATGTCCAGCCCTGCACCAGGACACGGTATTCGGCCACCTGTAACC 7113
PNRC2-ChRV     7035  ...............................................................................T 7114

PNRC2-BoRV     7114  TAATCCGTGTCAAATTTGAGGAGGCCGCAAAAACTGACCCCCGGTGGACAACCGGACTAATTTGGGGTCTAAATTTATAC 7193
PNRC2-ChRV     7115  ...................................................................C............ 7194

PNRC2-BoRV     7194  CAAACTCCGGCAGCTGGACTCCCTATCCAAATTAGACTACTAGTTAACCCGGTCTCAGCCTCGGTCCCGGTAGGGCCAAA 7273
PNRC2-ChRV     7195  ................................................................................ 7274

PNRC2-BoRV     7274  CCCGGTTCTAACAGGGAGAGCACCTTCTCAGTGAGGGAGCCGGCAGAAAGTCCCGACCACCGTTTCCCCATCCAAATCCC 7353
PNRC2-ChRV     7275  ................................................................................ 7354

PNRC2-BoRV     7354  CATCCCCATCGGCGCTCCCGGGGCCTACTCGCCTGCCTCCCGACCCGGAAACAAGCAATAGACTCTTCAACCTCATCAGA 7433
PNRC2-ChRV     7355  ................................................................................ 7434
```

```
PNRC2-BoRV      7434   GGCGCTTACCTCGCCCTGAACCAGACAAAGCCTGAATCTACCACCTCCTGCTGGCTCTGCCTGGCCACAGGCCCCCCTTA 7513
PNRC2-ChRV      7435   ......................................C......................................... 7514

PNRC2-BoRV      7514   CTATGAAGGTATTGCCTCTGTTAGTAATCTTACTAACTCCACTAGTCATTCTGGATGTGCATGGGACCAGCACAAGAAAC 7593
PNRC2-ChRV      7515   ................................................................................ 7594

PNRC2-BoRV      7594   TTACCCTAGCAGAGGTGTCAGGGTCGGGAACCTGTATAGGCCGGGTGCCCCCCAGTCACCAACATCTCTGTAATAGAACC 7673
PNRC2-ChRV      7595   ................................................................................ 7674

PNRC2-BoRV      7674   CTGGCAGTACCCAGAACTAGTCACTATCTAATACCCTCCGGACCAGACTGGTGGGCTTGCAAAACCGGACTTACCCCTTG 7753
PNRC2-ChRV      7675   ................................................................................ 7754

PNRC2-BoRV      7754   TATATCCGCAGCTGTCTTCAACGACAGTGAAGATTATTGTATATTAGTACAAGTTGTGCCCCGAGTTTATTATCAAACTG 7833
PNRC2-ChRV      7755   .G.............................................................................. 7834

PNRC2-BoRV      7834   GAGAGTCTTTTGAATCCCAGTTTGAGCAAAAATACCTCACTAGAATAAAGAGAGAACCTGTTTCCCTCACCCTCGCTGTT 7913
PNRC2-ChRV      7835   ................................................................................ 7914

PNRC2-BoRV      7914   ATGCTAGGATTAGGAGTAGCGGCTGGGGTCGGGACAGGAACCGCGGCATTAGTGCGTGGCAGCTACCACCTACAACAACT 7993
PNRC2-ChRV      7915   ................................................................................ 7994

PNRC2-BoRV      7994   CAGGGCAGCTGTAGATGAAGACCTCAGGGCCATAGAACACTCCATTACCAAACTTGAAGAATCTCTAACCTCCCTGTGCG 8073
PNRC2-ChRV      7995   .........C...................................................................C.. 8074

PNRC2-BoRV      8074   AAGTAGTACTCCAAAATCGACGGGGACTAGACATAATTTTTCTAAAAGAGGGCGGGCTCTGTGCAGCCCTTAAAGAGCAG 8153
PNRC2-ChRV      8075   ................................................................................ 8154

PNRC2-BoRV      8154   TGTTGTTTTTACGCTGATCATTCAGGAGTAGTTAAAGACTCTATGGCAAAACTTAGAAAAAGACTAGATGATAGACAAAA 8233
PNRC2-ChRV      8155   ................................................................................ 8234

PNRC2-BoRV      8234   AGAGAGAGAATCCCAACAAAGCTGGTTTGAAACTTGGTACAACCAATCCCCCTGGTTTAGTACTCTCATTTCCACCATCC 8313
PNRC2-ChRV      8235   ................................................................................ 8314

PNRC2-BoRV      8314   TAGGGCCCCTGATCCTGCTTATGCTCATTTTAACTTTTGGGCCCTGCATTTTTAACCGTTTGCTTACCCTAATTAAAGAT 8393
PNRC2-ChRV      8315   ................................................................................ 8394
                                                                                              env ←┐
PNRC2-BoRV      8394   AAATTAAACATAGTGCATGCTATGATCCTGACTCAACAGTACCAGGCAGTCAAGACTGACGAAGAGACTCAAGATTGAGC 8473
PNRC2-ChRV      8395   ...........................................................T.................... 8474
                                                ┌→ 3'LTR
PNRC2-BoRV      8474   CTCTAAGTCACAAAAAGAGGAGGGAATGAGAAACCATAGAGAAAATACAGCCACTAAAACACCAGATGGCCAGGAGTCAG 8553
PNRC2-ChRV      8475   ................................................................................ 8554

PNRC2-BoRV      8554   GGTGTGTCAATGGCTATCATAAACACCAGATGCCCAGGGTAAGACAGGGAG-CCCCCCCGCCCCTTGTTCCCGGAACTAA 8632
PNRC2-ChRV      8555   ...................................................C............................ 8634

PNRC2-BoRV      8633   AGCAGAAAAACATTTCCAGGAAATACCTCCTCCCGCCTATAAGCCCCTGACCAACCAGTGTGAGACAGCTAGCTCAGACC 8712
PNRC2-ChRV      8635   ................................................................................ 8714

PNRC2-BoRV      8713   ATAATTAGAACCAATCAGTTACTTCCAAACCTCGCGCCCAAAAACTGTTCAAATGTGTAACCCAATCAGCTTATTGTAAC 8792
PNRC2-ChRV      8715   ................................................................................ 8794

PNRC2-BoRV      8793   CTCCCCGTTTGAATTTGTGGTTTTTGCCTTTATAAGCAGTCTGTAACAATCATTCGGGGTCTCCTGGCCTTATGTGCTGG 8872
PNRC2-ChRV      8795   ...........-.................................................................... 8873

PNRC2-BoRV      8873   GGACCCTAGCGCGCTAGTAATAAATAGTGTCTCTTTGCTGTGATCTCTGTGTCGAGTGGTCTCTGGCGGCGGCCCCGTCC 8952
PNRC2-ChRV      8874   ................................................................................ 8953
                                             3' LTR ←─┬─→ 3' flanking
PNRC2-BoRV      8953   TGAGGCAGGAATCTTGAGTAAGGTTCCAACAGTGACATGTGTGGAAAGATTGGCAAATTCTGCTGGGAGTTTAAAAACAT 9032
PNRC2-ChRV      8954   C............................................................................... 9033
                                                 Target site duplication
PNRC2-BoRV      9033   GGTCTCTTTAGAAAATCCGCCTGAAGCCAATTTTTTGATATGACTCAACATTTAACACTTCCTTGATTTTATATGCGAGA 9112
PNRC2-ChRV      9034   .........................................................................C...... 9113

PNRC2-BoRV      9113   GAATCTGAAGGTCTTTCCCAGGAAAAAGATAAAGATAAAACTTACTGCGGGACTTGGAGACACTTACAAAAATTCCCCAG 9192
PNRC2-ChRV      9114   ............................................T................................... 9193

PNRC2-BoRV      9193   GAGTGACTGGATCTCAAGGGTGA--AGTTTTTTTAAATGGTCGTGGTTATAAGGTCAAAGAAAGCCATCTTACGAGAAAT 9270
PNRC2-ChRV      9194   .......................AG....................................................... 9273

PNRC2-BoRV      9271   TTTCTACCTTTTAGTTCAGGAAGTAGCCTTCTTTACAGTCAACGATTTTCGTAGTAAATCTAAATGATCACTTTCTCAAC 9350
PNRC2-ChRV      9274   ................................................................................ 9353

PNRC2-BoRV      9351   GTGTTTTTGAAATAATAAAGTAATCTTTACTTTTCTAGCTAGGACTTCTCTCAAACTTGTGTGCTGAGGAGACTCAGATG 9430
PNRC2-ChRV      9354   ...............-----............................................................ 9428

PNRC2-BoRV      9431   TTGGCCTCAGCTCCTAGGCTGAACTCAGCAGATCGGCCCATGAAAACTTCTGTATTGAGACAAAGGAAGGGATCTGTCAG 9510
PNRC2-ChRV      9429   ................................................................................ 9508

PNRC2-BoRV      9511   AAAGCAACACTTGTTATCTTGGGCTTGGCAGCAAGGAAGAGGACAGGTAGTGGAGATCCTACAATCTGAAAAGCAGACTG 9590
PNRC2-ChRV      9509   ............................................................G................... 9588

PNRC2-BoRV      9591   AAAGGTAATCATCATAATGGGCAAACATTTTATGATGGAACCATCCTAAACCAATTGTTTTTATTTTAAAATAGATTATG 9670
PNRC2-ChRV      9589   ................................................................................ 9668

PNRC2-BoRV      9671   CAGATTTGTGGGTAAGCTCATTTTTCTTAGGGAAGGTATAACAATAAAGGTTGTGCATATATTGAAGTCTGAACGGCTTG 9750
PNRC2-ChRV      9669   ................................................................................ 9748

PNRC2-BoRV      9751   CTTTTTTGAGTTTAAAGCAGCTGAATAGCCTGTTATTGAGTCTTGTCCAAAATATCTGGAGTTATAAGTTGGAGCCATAG 9830
PNRC2-ChRV      9749   ................................................................................ 9828

PNRC2-BoRV      9831   AAGCTTCTTAAAGATTTTGAACTTTTCTTAATGGAATTTTACTCAATATTTGTTTCCATTCACTTGTAGGTGACAAAGAA 9910
PNRC2-ChRV      9829   ................................................................................ 9908
```

```
PNRC2-BoRV     9911  GCTGAAGATGGGTGGTGGAGAGAGGTATAACATTCCAGCCCCTCAATCTAGAAATGTTAGCAAGAACCAACAACAGCTTA 9990
PNRC2-ChRV     9909  ............................................................T................... 9988

PNRC2-BoRV     9991  ACAGACAGAAGACCAAGGAACAGAATTCCCAGATGAAGATTGTTCATAAGAAAAAAGAAAGAGGACATGGTTATAACTCA 10070
PNRC2-ChRV     9989  ................................................................................ 10068

PNRC2-BoRV     10071 TCAGCAGCTGCCTGGCAGGCCATGCAAAATGGGGGGAAGAACAAAAATTTTCCAAATAATCAAAGTTGGAATTCTAGCTT 10150
PNRC2-ChRV     10069 ................................................................................ 10148

PNRC2-BoRV     10151 ATCAGGTCCCAGCTTACTTTTTAAATCTCAAGCTAATCAGAACTATGCTGGTGCCAAATTTAGTGAGCCGCCATCACCAA 10230
PNRC2-ChRV     10149 ................................................................................ 10228

PNRC2-BoRV     10231 GTGTTCTTCCCAAACCACCAAGCCACTGGGTCCCTGTTTCCTTTAATCCTTCAGATAAGGAAATAATGACATTTCAACTT 10310
PNRC2-ChRV     10229 ................................................................................ 10308

PNRC2-BoRV     10311 AAAACCTTACTTAAAGTACAGGTATAAAATAAGACAAATGTTTAAATTTAGTTATGTTCACGGATAGTTGTCAATTGGTC 10390
PNRC2-ChRV     10309 ................................................................................ 10388

PNRC2-BoRV     10391 TGAAACAAATTCGCTAGGGAATCTATTTGTGTAGAACTAATTAATGTAAA 10440
PNRC2-ChRV     10389 ...........T...................................... 10438
```

Figure S13. Alignment of the nucleotide sequence between PNRC2-BoRV and PNRC2-ChRV. The proviral sequences with the flanking sequences, 5’ and 3’ LTRs, *gag*, *pol*, and *env* genes are shown. Flanking 4-bp target site duplication (TSD) sequences are shown in gray. Dots indicate identical nucleotide sequences, and dashes indicate the missing nucleotides. Nucleotide sequences encoding Env are highlighted in green.

```
                                 5' flanking
CmRV1_2               1     CTGGTCTTGATTTCCTTGACCTCAGGCGATCCACCACCTCCGCCTCCTCAGAAGTGTTGGGATTACAGGTGTGAACCACT 80
Rhesus macaque chr14  1     ................................................................................ 80

CmRV1_2               81    GCACCCTGCCCTGAAGTACTTTTCTAAAGCTTTTTATTCTTTGTTACCAGAAAGTCTTCATTTGCATAGACCAATTCTCT 160
Rhesus macaque chr14  81    ................................................................................ 160

CmRV1_2               161   GATCTGCTGGCTCCAGAATTCAATACTTGTACTCCTTAGAGCAACTGACACCCTCTC------TCTTTTTTTTTTTTTTT 234
Rhesus macaque chr14  161   .........................................................TTTTTT.T............... 240

CmRV1_2               235   GAGACGGAGTCTTGCTCTGTAGCCCAGGCTGGAGTGCAGTGGCCGGATCTCAGCTCACTGCAAGCTCCGCCTCCCGGGTT 314
Rhesus macaque chr14  241   .........................C...................................................... 320

CmRV1_2               315   TACGCCATTCTCCTGCCTCAGCCTCCGGAGTAGCTGGGACTACAGGCGCCCGCCACCTCGCCCGGCTAGTTTTTTTTGTA 394
Rhesus macaque chr14  321   ................................................................................ 400

CmRV1_2               395   TTTTTAGTAGAGACGGGGTTTCACGGTGTTAGCCAGGATGGTCTCGATCTCTTGACCTCGTGATCCGCCCGTCTCGGCCT 474
Rhesus macaque chr14  401   ................................................................................ 480

CmRV1_2               475   CCCAAAGTGCTGGGATTACAGGTTTGAGCCACCGCGCCCGGCCCAACTGACACCCTCTCTACACAAAAATGGCAGAAGAG 554
Rhesus macaque chr14  481   ................................................................................ 560

CmRV1_2               555   TCCAAATAGTCTTTCTATTAAGCACCTTGTTCCTTTGGTTTTAAGTTTCTTGCTATGTGTTACTCCTGGAGAAGTATAGG 634
Rhesus macaque chr14  561   ................................................................................ 640

CmRV1_2               635   AGACTACAAACACCACAAAAAAACAATGATTCCAGGACTTCAGAGTTCTTCCATGGCCATTTCAGGTCTGAAATTCTGAA 714
Rhesus macaque chr14  641   ..................................A............................................. 720

CmRV1_2               715   CTCTCCTCTCAAATGGAAACAAAAAGGCCCATTCAACTTCAGTCCTCAGCATTTTACCCCTCTGGCAGGGGCCATCAAAT 794
Rhesus macaque chr14  721   ................................................................................ 800

CmRV1_2               795   ACCGTCTCTTCTTCCAACACCACATCATCTCCATCTCCCAAGGGATGACCATCCGACATCCTTAAATCCTAACCTTAAAT 874
Rhesus macaque chr14  801   ................................................................................ 880

CmRV1_2               875   GCCAATAGTTGTTGGTTTACAGTCTTTTACAAAAGTGAAAAAGCTGCAGTCCTTTAAGCAAATGAAACTTAAGGCAATAA 954
Rhesus macaque chr14  881   .................................................................G.....C........ 960

CmRV1_2               955   ATGATCCCATCCACTCTCAGGGAGGTACCCTCTGCTTCCCGTCTGTGTTCTCTAGAAACAGATTAACCAGTGCTTGTTAA 1034
Rhesus macaque chr14  961   ................................................................................ 1040

CmRV1_2               1035  TAAAGCATATTTCTCAGACTTGGCTCAAAATTTGCATTTTAGCTGTATCTATAAGGCTTCTGGGGCCAGGTGCTGTGGCT 1114
Rhesus macaque chr14  1041  ................................................................................ 1120

CmRV1_2               1115  CATGCCTATAATCCCAGCACTTTGGGAGGCCGAGGCAGGCAGATCACCTGAGGTCAGGTGTTTGAGACCAGACTTGCAAA 1194
Rhesus macaque chr14  1121  ................................................................................ 1200

CmRV1_2               1195  CATGGTGAAAAGCCACTTCTACTGGAAGTACAAAAGTTAGCCAGGTGTAGTGGCAGGCGCCTGTAATCCCAGCTACACGG 1274
Rhesus macaque chr14  1201  ................................................................................ 1280

CmRV1_2               1275  GAGACTGAGGTGGGAGAATCGCTTAAACCCGGGAGACAGAGGTTGCGGGGAGCCGAGATTGTGCCACTGCACTCCAGCAT 1354
Rhesus macaque chr14  1281  ................................................................................ 1360

CmRV1_2               1355  GGGTCACAGAGTGAGACTCCATCTCGGAAAAAAAGAAAGACTTTTGGGAATAAAACGTCAATATTCTCTAGGACCATTGC 1434
Rhesus macaque chr14  1361  ................................................................................ 1440

CmRV1_2               1435  TTAAGTTTAAGATGAAACTGCTTTATTATTCAAATCTGGAAATTTGAGAAAGGAGGAGTTACCGATAAGCATACTGGGAA 1514
Rhesus macaque chr14  1441  ................................................................................ 1520

CmRV1_2               1515  AACAGGGGTAAATCAGGTTGCCCCAGGCAAACTGGAACTTGTGGTTACTCTAATTGTAACTAGTCTTCTGAATGAGTGCT 1594
Rhesus macaque chr14  1521  ................................................................................ 1600

CmRV1_2               1595  GTTAGAAGAAAAAACTTTAGACAAATTTAACAGTGTTTAACAGAGCAAAGAACAATTCTAGAACTGGACAGTGCCCTGAA 1674
Rhesus macaque chr14  1601  ...............................-.CA............................................. 1679

CmRV1_2               1675  CCGGAAAAAGTTCAGAGTAACTCTAGGGCTGCTACCTGATCAGATAATGTTTATTTATTTATTATTTATTATGTTTTTTT 1754
Rhesus macaque chr14  1680  ..A............................................................................. 1759

CmRV1_2               1755  TCTGAGATGGAGTCTTGCTCTGTTGCCCAGACTAGAGTTTATTGGCTCAATCTCAGCTCACTGCAACTTGTACCTCCTGG 1834
Rhesus macaque chr14  1760  ................................................................................ 1839

CmRV1_2               1835  GTTCAAGCAATTCTCCTGCCTCAGCCTCCTGAGTAGCTAGAATTACAGGCGTGTGCCACCATGCCTGGCTAACTTTAGTA 1914
Rhesus macaque chr14  1840  ................................................................................ 1919

CmRV1_2               1915  TTTTTAGTAGAGACGGGGTTTCACCACGTTGCCAGGCTGGTCTCGAACGCCTGACCTCAGGTGATCCACCCGCCTCAGCC 1994
Rhesus macaque chr14  1920  ................................................................................ 1999

CmRV1_2               1995  TCGCAAAGTGCTGGGATTACAGGCGTGAGCCACTGCATCTCAGCCACTGGTGAGCTGAGACTGCACTGCTGCACTCTAGC 2074
Rhesus macaque chr14  2000  ....................................G.......................................C... 2079

CmRV1_2               2075  CTGGGTAACAAAAAAAAAAAAAAAAAGAGTAGATAACAGCCTATTTACATGTCCAGTTAGTTATGGTCTAATATGTATGG 2154
Rhesus macaque chr14  2080  ..........................-..................................................... 2158

CmRV1_2               2155  AGAACTCTTTGGACTGAACTTAAAGTATACAAGGAGGCAGCTTTAAGCTAAACTTAGCAGTGCTAAAACTCCCATTTTTC 2234
Rhesus macaque chr14  2159  ................................................................................ 2238

CmRV1_2               2235  CCCTCAGTTTTCCCTATCTTTGTCAGCTCAGCTCCAGTGCCACCCCCAGGAGTCTCGCTCTGTTGCCCAGGCTGGAATGC 2314
Rhesus macaque chr14  2239  ........................................................A....................... 2318

CmRV1_2               2315  AATGGTGCAATCTTGGCTCACTGCAACCTTCGCCTCCCGGGTTCAAGCGAGTCTCCTGCCTCAGGCTCCCGAGTAGCTGG 2394
Rhesus macaque chr14  2319  ................................................................................ 2398

CmRV1_2               2395  GATTATAGCCGCCCGCCACCATACCCAGCTACTTTTTGTATTTTTAGTACTGGCAGGGTTTCACCATGTTGGCCAGGCTA 2474
Rhesus macaque chr14  2399  ................................................................................ 2478
```

```
CmRV1_2               2475  GTCTCGAACTCCTGACCTCAGATGATCCACCCACCTCAGCTTTCCAAGGTGCTGGGATTGCAGGGGTGAGCCACCTTGCC 2554
Rhesus macaque chr14  2479  ................................................................................ 2558

CmRV1_2               2555  ACCTGGCCAGAAAAGTCTTTGGAGGCAGTTTTCTCCCTAATTATTCATCTTTCTCAATTAGGATTTCTCTAAACCTGAAT 2634
Rhesus macaque chr14  2559  ......G....G.................................................................... 2638

CmRV1_2               2635  CTCCTTAAGAGAGAGCTTAAATGTTTGATTTAGAGTGCTGTTTTAGAAAGACCTTGGGCAAATTACTTAATCTGTCTGTG 2714
Rhesus macaque chr14  2639  ...................................G............................................ 2718

CmRV1_2               2715  GTGGTTTCCTCACTTGTAAAGCAGGTTCTACTATAATACTAACAGCACTTACTTTATAGGGTATTGTGAGAGTAAAATCA 2794
Rhesus macaque chr14  2719  .......................................................................A........ 2798

CmRV1_2               2795  GTTCCTATGTGCAAAGTACTCAGAGAACCACCTGACCTAAGTGCTCGATAATGTTAGCTATTACCTCCTCTAAAAATGTT 2874
Rhesus macaque chr14  2799  ...................................T............................................ 2878

CmRV1_2               2875  GTATTTTATTTTATTTTATTTACTTTTATAGATGGGGTCTGTCTCTGTCATCCAGGCTGGAGTGCGGTGGTACAATCTTG 2954
Rhesus macaque chr14  2879  ............................................T...................T............... 2958
                 5’ flanking ←─┴─→ 5’ LTR
CmRV1_2               2955  GCTTGAAACTAGGCCTCATAAAACTTAGAAACTAGGCCTCATATAGAAAAAATTAACACCAGGTGGCTCTGGATAGGGTC 3034
Rhesus macaque chr14  2959  ................................................................................ 3038
                        Target site duplication
CmRV1_2               3035  CCACCCTGCCTCGATAGGGACCCACCCTGATAGGGTCCCACCCTGCCAATTCCGGGAAACAACCTCATGGGGTCCCACCC 3114
Rhesus macaque chr14  3039  ............A................................................................... 3118

CmRV1_2               3115  TGCCAATTCCGGGGGTCCCACCCTGCCTCGAAGTTCCCGGAATCAACAACTCCAGGAAAAAAACCTCATAAGGTCCTGCT 3194
Rhesus macaque chr14  3119  ................................................................................ 3198

CmRV1_2               3195  CTAACCAATTAGCATAAGACACCTTGCTCAGGCCATAGCTAGACCCAATCATTTTGCGCCTTAAGCTTTGTTTGAATTTC 3274
Rhesus macaque chr14  3199  ................................................................................ 3278

CmRV1_2               3275  GCGCCATAAGCTGTGTTTGAACTTGTGTTTGCCTATATAAACAGCCTGTAACAAGCAGTCGGGGTCCCAGGGCCAACTTA 3354
Rhesus macaque chr14  3279  .....G.......................................................................... 3358

CmRV1_2               3355  GAGCTTGGGACCCTAGTGCGCTAGTAATAAATAACTCTCTGCTGTGAATCTCGTGTCGGTGATCCTTCGCGGCGACCCCT 3434
Rhesus macaque chr14  3359  ................C........................................................A...... 3438
                                                     5’ LTR ←┐
CmRV1_2               3435  GCCCAGGAGGGAATCGACAGTTCGGTTCCAACATTTGGTGCATTGGCCGGGAAGTGGGGTCGTCCGAGGACCCCCGACCC 3514
Rhesus macaque chr14  3439  ................................................................................ 3518

CmRV1_2               3515  ATCCGGCGGAGACCCATCTGGCCCGGGCCACGGACTGCTGACTGAACGGACCCACCAGGTACTTTCGTTTTGTTCTGTCT 3594
Rhesus macaque chr14  3519  ..............................T................................................. 3598

CmRV1_2               3595  GTCTTGCCGGCTAACTCTGAACTCTGGGGAGTACTCCTTCTGAATTAAGTGGGGAAGGGGGACAGACGTGTCCGGCACCT 3674
Rhesus macaque chr14  3599  ........A....................................................................... 3678

CmRV1_2               3675  TCCCACTTACGCCCCGGGGGACGCCCTGGCAGTAGTCTGGGAGAAGGCTGACGACTCAGTCAGCCTCCTCAAATCTGTAG 3754
Rhesus macaque chr14  3679  ................................................................................ 3758

CmRV1_2               3755  GCAGGTCGCCCCTGCCGTCTGAATATTTTGTGATCTTGTGGCACCACTCTCTGGCCGCGCGGCTTCTCCTTACTTGTCTG 3834
Rhesus macaque chr14  3759  ................................................................................ 3838
                                                                           ┌→ gag-pol
CmRV1_2               3835  GTCTTTATTTTTGTTACTTTCGTTTTGTCCTTGTTATACGTGGACGAAATGGGACAGACGTTGACGACTCCTTTGTCTCT 3914
Rhesus macaque chr14  3839  ............................................T................................... 3918

CmRV1_2               3915  AACCCTGACTCACTTCCCTGACGTCCGGGCTCGAGCCCACCACCTCTCTGTAGAAGTCCGTAAGGGACGATGGAAAACAT 3994
Rhesus macaque chr14  3919  ................................................................................ 3998

CmRV1_2               3995  TCTGCTCGTCCGAATGGCCAACCCTCCATGGGGAATGGCCCCGGGACGGAACATTTAACCTCTCAATTATCTTGCAGGTT 4074
Rhesus macaque chr14  3999  ..............................................T................................. 4078

CmRV1_2               4075  AAAGCAAAAGTGATGGATCCTGGGCCACTCGGACACCAGGACCAGGTGGCCTACATAATTATTTGGGAGGATCTGGTCCA 4154
Rhesus macaque chr14  4079  ................................................................................ 4158

CmRV1_2               4155  AAATCCTCCCTCTTGGGTGAAACCCTTCCTCCATTCCCCTTCCCCATCCCAATCTACCCTCCTTGCCTTAGAAGCCCCAA 4234
Rhesus macaque chr14  4159  ................................................................................ 4238

CmRV1_2               4235  AGAATCGGAATCTGGACCCGCCTAAGCCAGTCCTCCCAGATGAACCCCAGAGGGATCTCCTCCTTCTTGACCCCCTGCCT 4314
Rhesus macaque chr14  4239  ................................................................................ 4318

CmRV1_2               4315  CCTCCACCTCAGAACCCCCTTCTGGGACCTCCACCTTACACTTCACCCTTGCCCCCTGTCTTGTCCCCAGCTCTTTCCTC 4394
Rhesus macaque chr14  4319  ................................................................................ 4398

CmRV1_2               4395  TACCGCCTCGGCCCCTACCCTTTCTCCAACTTCTCCCTCGGCCCCTACCCTTTCTTTTTTTTTTTTTTTTTTTTTTTTGA 4474
Rhesus macaque chr14  4399  .......................................A..............-......................... 4477

CmRV1_2               4475  GACGGAGTCTGGCTCTGTCGCCCAGGCTGGAGTGCAGTGGCCAGATCTCAGCTCACTGCAAGCTCCGCCTCCCGGGTTTG 4554
Rhesus macaque chr14  4478  ................................................................................ 4557

CmRV1_2               4555  CGCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGAGACTACAGGCGCCCGCCACCTCGCCCGGCTATTTTTTTGTATTTT 4634
Rhesus macaque chr14  4558  ..................................................................G............. 4637

CmRV1_2               4635  TTTTAGTAGAGACGGGATTTCACTGTGTTAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCCGCCCGTGTCGGCCTC 4714
Rhesus macaque chr14  4638  ......................................T......................................... 4717

CmRV1_2               4715  CCAAAGTGCTGGGATTACAGACTTGAGCCACCACGCCCGGCCCCTCCCTCCACCCCGTCTCCTTCTCCAGCCCCGCCCAA 4794
Rhesus macaque chr14  4718  .................................T.............................................. 4797

CmRV1_2               4795  GCTCACCCCTCGGACGCCGCCGCCAACACCTCCACGTCTCCGCTTGCGGCGGACTGAGGACCCAGATGGCCCTTCCACTT 4874
Rhesus macaque chr14  4798  ................................................................................ 4877

CmRV1_2               4875  GGCAATCCTCCCTTTTTCCCCTCCGTACTGTCAATCGCACGGTCCAGTACTGGCCCTTCTCTGCCTCTGACCTCTACAAC 4954
Rhesus macaque chr14  4878  ................................................................................ 4957
```

```
CmRV1_2               4955  TGGAAAACCCATAACCCTTCCTTTTCCCAAGACCCCCAGGCCCTAACCTCGTTGATAGAATCCATTCTCCTCACTCACCA 5034
Rhesus macaque chr14  4958  ....................................................A................T.......... 5037

CmRV1_2               5035  GCCCACTTGGGATGATTGCCAGCAACTCTTGCAGGTCCTCCTAACCACTGAAGAAAGGCAGCGAGTCCTCCTGGAGGCCC 5114
Rhesus macaque chr14  5038  ................................................................................ 5117

CmRV1_2               5115  GGAAAAATGTGCCAGGACCAGGAGGCCTCCCAACCCAACTTCCCAATGAAATAGACGAGGGATTTCCCCTCACCCGCCCA 5194
Rhesus macaque chr14  5118  ................................................................................ 5197

CmRV1_2               5195  GACTGGGACTATGAAATGGCACCAGGTAGGGAGAGTCTCCAAATCTATCGCCAGGCTCTGTTGGCAGGTCTCAAAGGGGC 5274
Rhesus macaque chr14  5198  .................................................................G.............. 5277

CmRV1_2               5275  AGGAAAGCGCCCCACAAATTTGGCCAAGGTAAGGACCATAACTCAGGAAAGGGATGAAAGCCCGGCAGCCTTCATGGAAA 5354
Rhesus macaque chr14  5278  ...............................................C......C.......T................. 5357

CmRV1_2               5355  GGCTTCTGGAAGGGTTCCGAATGTATACCCCATTCGATCCAGAGGCCCTAGAACATAAGGCTACCGTAGCTATGGCATTC 5434
Rhesus macaque chr14  5358  ................................................................................ 5437

CmRV1_2               5435  ATAGACCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5514
Rhesus macaque chr14  5438  ................................................................................ 5517

CmRV1_2               5515  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5594
Rhesus macaque chr14  5518  ................................................................................ 5597

CmRV1_2               5595  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5674
Rhesus macaque chr14  5598  ................................................................................ 5677

CmRV1_2               5675  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5754
Rhesus macaque chr14  5678  ................................................................................ 5757

CmRV1_2               5755  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5834
Rhesus macaque chr14  5758  ................................................................................ 5837

CmRV1_2               5835  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5914
Rhesus macaque chr14  5838  ................................................................................ 5917

CmRV1_2               5915  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5994
Rhesus macaque chr14  5918  ................................................................................ 5997

CmRV1_2               5995  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6074
Rhesus macaque chr14  5998  ................................................................................ 6077

CmRV1_2               6075  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6154
Rhesus macaque chr14  6078  ................................................................................ 6157

CmRV1_2               6155  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6234
Rhesus macaque chr14  6158  ................................................................................ 6237

CmRV1_2               6235  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN----------- 6303
Rhesus macaque chr14  6238  .....................................................................CTAAGTATCAA 6317

CmRV1_2               6303  -------------------------------------------------------------------------------- 6303
Rhesus macaque chr14  6318  GCCCCAATCATAATTGATTTAAAGCCCACGGCGGTGCCCGTGTCTATCAAGCAATATACCATGAGCCGAGAGGCTCATAT 6397

CmRV1_2               6303  -------------------------------------------------------------------------------- 6303
Rhesus macaque chr14  6398  AGGAATTCGGCAGCACATTAACAAATTTCTAGAACTCGGAGTGTTGCGACCTTGTCGCTCGCCCTGGAACACTCCTCTTC 6477

CmRV1_2               6303  -------------------------------------------------------------------------------- 6303
Rhesus macaque chr14  6478  TGCCAGTAAAAAAGCCTGGTACTCAGGATTACAGGCCTGTCCAAGACTTGAGAGAAATTAACAAAAGAACCATGGACATC 6557

CmRV1_2               6303  -------------------------------------------------------------------------------- 6303
Rhesus macaque chr14  6558  CATCCCACGGTCCCCAATCCTTACAACTTACTCAGCACCTTAAAACCAGGCTATAACTGGTATACAGTATTAGATTTAAA 6637

CmRV1_2               6303  -------------------------------------------------------NNNNNNNNNNNNNNNNNNNNNNNNN 6328
Rhesus macaque chr14  6638  AGATGCTTTCTTCTGTTTACCTCTGGCCCTCCAAAGCCAAGAACTCTTTGCCTTT......................... 6717

CmRV1_2               6329  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6408
Rhesus macaque chr14  6718  ................................................................................ 6797

CmRV1_2               6409  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6488
Rhesus macaque chr14  6798  ................................................................................ 6877

CmRV1_2               6489  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6568
Rhesus macaque chr14  6878  ................................................................................ 6957

CmRV1_2               6569  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6648
Rhesus macaque chr14  6958  ................................................................................ 7037

CmRV1_2               6649  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6728
Rhesus macaque chr14  7038  ................................................................................ 7117

CmRV1_2               6729  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6808
Rhesus macaque chr14  7118  ................................................................................ 7197

CmRV1_2               6809  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6888
Rhesus macaque chr14  7198  ................................................................................ 7277

CmRV1_2               6889  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6968
Rhesus macaque chr14  7278  ................................................................................ 7357

CmRV1_2               6969  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 7048
Rhesus macaque chr14  7358  ................................................................................ 7437
```

```
CmRV1_2               7049  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 7128
Rhesus macaque chr14  7438  ................................................................................ 7517

CmRV1_2               7129  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 7208
Rhesus macaque chr14  7518  ................................................................................ 7597

CmRV1_2               7209  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 7288
Rhesus macaque chr14  7598  ................................................................................ 7677

CmRV1_2               7289  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 7368
Rhesus macaque chr14  7678  ................................................................................ 7757

CmRV1_2               7369  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 7448
Rhesus macaque chr14  7758  ................................................................................ 7837

CmRV1_2               7449  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 7528
Rhesus macaque chr14  7838  ................................................................................ 7917

CmRV1_2               7529  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 7608
Rhesus macaque chr14  7918  ................................................................................ 7997

CmRV1_2               7609  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 7688
Rhesus macaque chr14  7998  ................................................................................ 8077

CmRV1_2               7689  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 7768
Rhesus macaque chr14  8078  ................................................................................ 8157

CmRV1_2               7769  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 7848
Rhesus macaque chr14  8158  ................................................................................ 8237

CmRV1_2               7849  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGGCTATAGAAAACAAAGACACTAAAAACTGGGAAAAAGGAGGAAAAAT 7928
Rhesus macaque chr14  8238  ................................................................................ 8317

CmRV1_2               7929  AGTCCTTCCCCAAAAGGAGGCCCTGGCAATGATCCAGCAGATGCATGCCTAGACACACTTGAGTAATTGAAAGCTAAAT 8008
Rhesus macaque chr14  8318  ..................................................G....A........................ 8397

CmRV1_2               8009  TACTGATTGAAAAAACTGACTTTCTAATCCCAAAGGCAAGTACCCTCGTAGAACAAGTGACATCTGCCTGTAAGGTCTGT 8088
Rhesus macaque chr14  8398  ................................................................................ 8477

CmRV1_2               8089  CAGCAGGTAAACGCTGGGGCTACCCGAGTGCCAGAAAGGAAACGAACTCGTGATAACCGCCCAGGAGTCTATTAGGAAAT 8168
Rhesus macaque chr14  8478  .........................T..........G...............G....................G...... 8557

CmRV1_2               8169  AGACTTCACTGAAGTAAAACCTCACTATGCTGGATATAAGTACTTACTGGTGTTTGTAGATACCTTTTCAGGATGGGTAG 8248
Rhesus macaque chr14  8558  ................................................................................ 8637

CmRV1_2               8249  AAGCCTACCCCACCCGGCAAGAAACGGCACACATAGTAGCCAAGAAAATTTTGGAAGAAATCTTTCCTAGATTCGGACTT 8328
Rhesus macaque chr14  8638  ................................................................................ 8717

CmRV1_2               8329  CCCAAGGTAATTGGGTCAGATAACGGGCCGGCCTTCGTTTCTCAGGTAAGTCAGGGGCTCGCCAGGATATTGGGGATTAA 8408
Rhesus macaque chr14  8718  ................................................................................ 8797

CmRV1_2               8409  TTGGAAATTACATTGTGCCTATAGACCCCAGAGCTCAGGACAGGTAGAAAGAATGAATAGAACAATAAAAGAGACCCTTA 8488
Rhesus macaque chr14  8798  ................................................................................ 8877

CmRV1_2               8489  CTAAATTGACCTTAGAGACTGGTTTAAAAGATTGGAGACGCCTCCTATCCTTAGCTCTGTTAAGGGCCCGAAATACGCCT 8568
Rhesus macaque chr14  8878  ..............................................G................................. 8957

CmRV1_2               8569  AACCGTTTTGGGCTCACTCCATATGAAATCCTCTACGGAGGACCTCCCCCTTTGTCAACCTTGCTTAATTCCTTCTCCCC 8648
Rhesus macaque chr14  8958  ................................................................................ 9037

CmRV1_2               8649  CTCCGATCCTAAGACTGACCTACAGGCCCGGCTAAAAGGACTCCAAGCAGTACAGGCCCAAATCTGGGCCCCCTTGGCAG 8728
Rhesus macaque chr14  9038  ................................................................................ 9117

CmRV1_2               8729  AACTGTACCAACCAGGACATCCACAGACCAGTCACCCCTTCCAGGTGGGAGACTCTGTCTACGTTAGACGGCATCGCACT 8808
Rhesus macaque chr14  9118  ................................................................................ 9197

CmRV1_2               8809  CAAGGACTAGAGCCTCAGTGGAAAGGACCCTACATTGTTCTCCTGACCACGCCCACAGCCATAAAGGTTGACGGAATCTC 8888
Rhesus macaque chr14  9198  .........................................T...................................... 9277
                                                                                              → env
CmRV1_2               8889  CACTTGGATCCACGCATCCCACGCCAAGGCTGCTCCAGAGACGCCCGGACCAACTCCACCTAAGACATGGAGACTCCGAC 8968
Rhesus macaque chr14  9278  ................................................................................ 9357
                                                           gag-pol ←
CmRV1_2               8969  GCTCCGAGGACCCGCTCAAGATAAGACTCTCTCGTATCTAGCCCCTTGCTTATTACTAGCCCTCCTTCCCTGCGTCGCTG 9048
Rhesus macaque chr14  9358  ........................................................................T....... 9437

CmRV1_2               9049  ACAGTAATAACCCCCACCGGCCATATAACCTGACCTGGCAGGTAATTGATTTTAGTACTCATGAGGTCTTAGATAAAACC 9128
Rhesus macaque chr14  9438  ................................................................................ 9517

CmRV1_2               9129  TCAAAAATTGCTCCTATGGGGACTTGGTTCCCTGACCTCTATTTCAACCTAGACAAAATAGCAAAGATAGATGACATGGA 9208
Rhesus macaque chr14  9518  ................................................................................ 9597

CmRV1_2               9209  AGGGGGAGAATGGAGAAAACAGGCTAGAAGGGTGTCCGTAAGCCGGAACGGGTTCTATGCCTGTCCCGGATTCAGGACAG 9288
Rhesus macaque chr14  9598  ................................................................................ 9677

CmRV1_2               9289  GAGAAATGAGAAAAACTTGTGGAGAAATAGATGCCCTGTTTTGCGCTAGTTGGTCCTGTATAACTACTAATGATGGAGAA 9368
Rhesus macaque chr14  9678  ................................................................................ 9757

CmRV1_2               9369  TGGAAATGGGCCACGAAACCCTGGTACATAACCATGTCCTTTGTCCAGTGCTGCACCAGAACCCGATATTCAAAAACTTG 9448
Rhesus macaque chr14  9758  ................................................................................ 9837

CmRV1_2               9449  CAATCTGGTCCGCATCAAGTTTGAAGATGCAGCAAAATCTGATAACCGTTGGATATCCGGGTTAATATGGGGCCTATATT 9528
Rhesus macaque chr14  9838  ................................................................................ 9917
```

```
CmRV1_2               9529  TATACCAAAAGCCACTGTATGGAATCCCTATCCAAATCAAATTAATAGTCAACCCTATCACAGCCCCTGTCGCAGTAGGA 9608
Rhesus macaque chr14  9918  ................................................................................ 9997

CmRV1_2               9609  CCAAACCAAGTTTTATCAAAAACAAGGAAGCCCCTGGTTCCTGCACCAAGAGAGCCCCAACCAAGAGCTCCTAAAAGCAC 9688
Rhesus macaque chr14  9998  ................................................................................ 10077

CmRV1_2               9689  TTCCCCGCCTCTAATCTCCACCTCCTCTAAATACACACCCTCAGCCCAAAACGTCACCCGCGGGCCCCTTGACCTGGGAA 9768
Rhesus macaque chr14  10078 ............................................T................................... 10157

CmRV1_2               9769  TAGGTGACAGGCTCCTAAATCTCATAAAGGGCTCCTATTTCGCTTTAAACCAGACAAAGCCAGAATTTACCTCCTCTTGC 9848
Rhesus macaque chr14  10158 ................................................................................ 10237

CmRV1_2               9849  TGGCTATGTCTGGCAACAGGCCCCCCTTACTATGAAGGCATTGCCTCCACTAATAATTTCACTAACTCCGCCAATCCTAC 9928
Rhesus macaque chr14  10238 ................................................................................ 10317

CmRV1_2               9929  TGGATGCGCATGGGAACAACAAAGAAAACTAACCCTGGCTGAAGTTTCTGGGTCAGGAACCTGCATAGGTCAAGTGCCCC 10008
Rhesus macaque chr14  10318 ......................................................G..............C.......... 10397

CmRV1_2               10009 CTAGTCATCAGCATCTTTGTAATGTAACCTTGACAGTACCCAGCTCCAATCACTATTTGGTCCCCTCCGAGACGGACTGG 10088
Rhesus macaque chr14  10398 ................................................................................ 10477

CmRV1_2               10089 TGGGCTTGCAACACTGGGCTCACCCCCTGTGTATCCACAGCCGTTTTTAGCAGTGGCACCCACTATTGCGTGTTGGTACA 10168
Rhesus macaque chr14  10478 ................................................................................ 10557

CmRV1_2               10169 AGTTGTTCCCCGAGTATACTATCACTCTGGAGACTCCTTTGATCTCCGGTATGAGCAAAAAACTCATACTAGACCAAAGA 10248
Rhesus macaque chr14  10558 ................................................................................ 10637

CmRV1_2               10249 GAGAACCTATCTCCCTCACCCTCGCCGTAATGTTAGGAATTGGGGTAGCGGCTGGAGTTGGGACCGGGACAGCAGCCCTA 10328
Rhesus macaque chr14  10638 ................................................................................ 10717

CmRV1_2               10329 GTGCATGGTAACCATCATCTGCAACAACTTAGAGTAGCCATAGATGAAGACCTTAGAGCCATAGAACAATCCATCACAAA 10408
Rhesus macaque chr14  10718 ................................................................................ 10797

CmRV1_2               10409 ACTTGAAGAGTCCTTGACTTCTCTGTCTGAAGTTGTATTACAAAACCGACGAGGACTAGAAATTGTCTTTCTGAAAGAGG 10488
Rhesus macaque chr14  10798 ...............................................C................................ 10877

CmRV1_2               10489 GCGGGCTCTGTGCAGCCCTGAAAGAGCAATGTTGTTTTTATGCAGATCATTCAGGAGTAGTTAAAGATTCTATGGCAAAA 10568
Rhesus macaque chr14  10878 ................................................................................ 10957

CmRV1_2               10569 CTGAGAGAAAGATTAGACAAGAGGAAAAAAGAGAGAGAATCTCAGCAAAATTGGTTTGAAAATTGGTACAACCAATCCCC 10648
Rhesus macaque chr14  10958 ..A....................A........................................................ 11037

CmRV1_2               10649 TTGGCTTAGCACCCTAATCTCCACCATCTTAGGACCCCTCATCCTGCTCACGCTCATCCTGACTTTCGGGCCATGCATAC 10728
Rhesus macaque chr14  11038 ................................................................................ 11117

CmRV1_2               10729 TCAACCGCTTACTCACCCTTATTAAAAATAGATTAAACATAGTACATGCTATGGTTCTGACCCAACAATACCAGACCCTC 10808
Rhesus macaque chr14  11118 ................................................................................ 11197
                                                  env ←┐                            ┌→ 3'LTR
CmRV1_2               10809 AGGACTGAAGAAGAGGCTCAAGATTGAGCCTCTGACACAAAAAGAGGAGGGAATGAAACTAGGCCTCATAAAACTTAGAA 10888
Rhesus macaque chr14  11198 ................................................................................ 11277

CmRV1_2               10889 ACTAGGCCTCATATAG-AAAAATTAACACCAGGTGGCTCTGGATAGGGTCCCACCCTGCCTCGATAGGGACCCACCCTGA 10967
Rhesus macaque chr14  11278 ................A............................................................... 11357

CmRV1_2               10968 TAGGGTCCCACCCTGCCAATTCCGGGAAACAACCTCATGGGGTCCCACCCTGCCAATTCCGGGGGTCCCACCCTGCCTCG 11047
Rhesus macaque chr14  11358 ................................................................................ 11437

CmRV1_2               11048 AAGTTCCCGGAATCAACAACTCCAGGAAAAAAACCTCATAAGGTCCTGCTCTAACCAATTAGCATAAGACACCTTGCTCA 11127
Rhesus macaque chr14  11438 ................................................................................ 11517

CmRV1_2               11128 GGCCATAGCTAGACCCAATCATTTTGCGCCTTAAGCTTTGTTTGAATTTCGCGCCATAAGCTGTGTTTGAACTTGTGTTT 11207
Rhesus macaque chr14  11518 ................................................................................ 11597

CmRV1_2               11208 GCCTATATAAACAGCCTGTAACAAGCAGTCGGGGTCCCAGGGCCAACTTAGAGCTTGGGACCCTAGTGCGCTAGTAATAA 11287
Rhesus macaque chr14  11598 ................................................................................ 11677

CmRV1_2               11288 ATAACTCTCTGCTGTGAATCTCGTGTCGGTGATCCTTCACGGCGACCCCTGCCCAGGAGGGAATCGACAGTTCGGTTCCA 11367
Rhesus macaque chr14  11678 ......................................G......................................... 11757
                      3' LTR ←─┴─→ 3' flanking
CmRV1_2               11368 ACAGGCTCACTGTAGCCTTGAACTCCTGTGTTCAAGTGATCCCAGAACTTCAGCCTCCCAATTAAAAATAGTTCTCAAAA 11447
Rhesus macaque chr14  11758 ................................................................................ 11837
                         Target site duplication
CmRV1_2               11448 AAAAAAAATAGTTCTCAGGCCGGGCGCGGTGACTCACACCTGTAATCCCAGCACTTTGGAAGGCTGAGGTGGGTTCGGGG 11527
Rhesus macaque chr14  11838 .....................A.......................................................... 11917

CmRV1_2               11528 CAGAGGGGGGGAAATCACCTGAGGTTGGGAGTTTGAGACCATCCTGGCCAACATGGAGAAACCCCTCTCTACTAAAAATA 11607
Rhesus macaque chr14  11918 ................................................................................ 11997

CmRV1_2               11608 CAAAATTAACTGGGCGTGGTGGCGCATACCTGTAATCCCAGCTACTCTGGAGCCTGAGGCAGGAGAATTGCTTGAACCCG 11687
Rhesus macaque chr14  11998 ................................................................................ 12077

CmRV1_2               11688 GGAGGCGGAGGTTGCAGTGAGCTGAGATCGTGCCACTGCACTTCAGCCTGGGCAACAAGAGCGAAGCTCCATCTCAAAAA 11767
Rhesus macaque chr14  12078 .....T.......................................................................... 12157

CmRV1_2               11768 TAAAAAAAGAAATAAATAATAATAATAGTTCTCTTTTCTTCCAGAGTTAAGTACCAGATCAATAAAATTATTATTTTCTT 11847
Rhesus macaque chr14  12158 ................................................................................ 12237

CmRV1_2               11848 TTTAGAGACAGGGTCTCACCCTGTTACCCAGGCTGGAGTGCAGTGGCATAATCATAGCTCGCTGCCGCCTCGATTTCCTG 11927
Rhesus macaque chr14  12238 ................................................................................ 12317

CmRV1_2               11928 GGCTCAAGTGATCATCCTGCCTCCGCCTCTGGAGTAGCTGGGACTATAGGTGTGCACCACCCATGCCCAGCTGATTTTTG 12007
Rhesus macaque chr14  12318 ................................................................................ 12397
```

```
CmRV1_2               12008 TAGGTTTTGTAGAGATGGCATCTTGCTATATTGCCTGGACTGGTCTGGAACTCCTAGCCTGAAGTGATCCTCCCACCTCA 12087
Rhesus macaque chr14  12398 ..........G..................................................................... 12477

CmRV1_2               12088 ACCCCCCAAGGTGCCGGAATTATAAATGTCAGCCACGGAGAGTTCATATTCTGAATACCAAAGCCAAATAGAGTCCACAG 12167
Rhesus macaque chr14  12478 ................................................................................ 12557

CmRV1_2               12168 GGTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGGCGCGATCTC 12247
Rhesus macaque chr14  12558 ..-....G.............G.......................................................... 12636

CmRV1_2               12248 GGCTCACTGCAACCTTCGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTACAGGCACCT 12327
Rhesus macaque chr14  12637 ................................................................................ 12716

CmRV1_2               12328 GCCACCACGCCTGGCCAATTTTTTGTATTTTTAGTAGAGACAGGGTTTCACTGTGTTAGCCAGGATGGTCTCGATCTCCT 12407
Rhesus macaque chr14  12717 ................................................................................ 12796

CmRV1_2               12408 GACCTTGTGATCCACCCGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCACGCCCGGCCGAGAACATGG 12487
Rhesus macaque chr14  12797 .....C........................................................C................. 12876

CmRV1_2               12488 ATTTTAATTGCATGGGTCCACTTATATGCAGATCAAAATACACACGTGCATCAAAAATACTGTATTCAGGCTGGGTGCAG 12567
Rhesus macaque chr14  12877 ................................................................................ 12956

CmRV1_2               12568 TGGCTCACACTTGTAATCCCAACATTTTGGGAGGCCAAGGTGGGAGGATCGCTTGAGCTCAGGAATTCCAGGCTTCAGTG 12647
Rhesus macaque chr14  12957 ..........................................................................G..... 13036

CmRV1_2               12648 AGCTGTGATCGTGCCATTGCACCCCAGCCTGGGCAACAGAGAAATACCCTGTCTCAAAAACAAACAACACACAAACAGTA 12727
Rhesus macaque chr14  13037 ................................................................................ 13116

CmRV1_2               12728 TTCAAGGGTTATGAAACCTGAGTATGTAGAGTTTTCCTATAAGTGGGTTCTGCAGGGTCAACTTCACAATTTGAGTATGT 12807
Rhesus macaque chr14  13117 ................................................................................ 13196

CmRV1_2               12808 GCAGACTTCATTATGGTAGAGCCAGGTGGAGGTCCTGGAATCAATCCCCTGTGTATACGTGTTAACTGTACTTCTTTTTT 12887
Rhesus macaque chr14  13197 ...........................................................C.................... 13276

CmRV1_2               12888 TTTTCTTCTTCTTCCTTTTTGAGACACAGTCTTGCTCTGTCACCCAGGATGGAGTGCAGTGGCGCAATCTTGGCTCACTG 12967
Rhesus macaque chr14  13277 ................................................................................ 13356

CmRV1_2               12968 CAACCTCCTCCACCTCCCAGGTTCAAGTGATTCTCCTGCCTCAAAAAAAATAAATAAATAAAAATAAAATAGTAAACAAA 13047
Rhesus macaque chr14  13357 ................................................................................ 13436

CmRV1_2               13048 AGAAAACAGAAGTGAGGGAAGTAAGGTACAGAAATAGCTGGATTGGTTGCAGCTTGGCATTTGCCTTATTTGAACATGAT 13127
Rhesus macaque chr14  13437 ................................................................................ 13516

CmRV1_2               13128 TTGAACAGTTGTCTGCCTGTGATTTCTGGTTGGCTGGGACTATAGGCACACGCCACCACACCCGGCTAATTTTTGTATTT 13207
Rhesus macaque chr14  13517 ..........................................................TG.................... 13596

CmRV1_2               13208 TTAGTAGAGACGGAGTTTCACCATGTTGGCCAGGATGGTCTTGATCTCCTGACCTCGTGATCCACCCACCTTGGCCTCCC 13287
Rhesus macaque chr14  13597 ................................................................................ 13676

CmRV1_2               13288 AAAGTGCTAGGATTACTGGCGTGAGCCACCACACCAGCCTAGTTAACTGTACTTCTAAAATGAACTCATTAACTTTACCA 13367
Rhesus macaque chr14  13677 ................................................................................ 13756

CmRV1_2               13368 AAAGCTATTTCAATTCTCCACACACTACATTGCCTACATTACACATTTGAGGCAGATTGGAAATCCATTTATTTTTCAAC 13447
Rhesus macaque chr14  13757 ................................................................................ 13836

CmRV1_2               13448 CAACCATAGATCCTTTTCTGAAACATTTTCTTAACACCTATAATTTAACCTTAAATAAGTAAACTCTTTAATCTAGGGGT 13527
Rhesus macaque chr14  13837 ................................................................................ 13916

CmRV1_2               13528 TATTACTGAAATACCTGTATTTTATCCCTCAAAGGCCATTTTCCTAGCTGCTGAGGAGCTGGCCCTTAGGGACATTTCTG 13607
Rhesus macaque chr14  13917 ................................................................................ 13996

CmRV1_2               13608 ACTAGATCATTCCTGGAAAGGAATCCATAGCTTAAAGAGCAGTCACACCTGTCAATCAGTCATAGCCTGCGGCTTTCAAA 13687
Rhesus macaque chr14  13997 ................................................................................ 14076

CmRV1_2               13688 TCCAGTTGCTCAGAAGCCACTGGCAAAAAACATTTAGAGTGAAGCAGAGAAGAAACATGTTTCTTTATTTCTTCTTTTCT 13767
Rhesus macaque chr14  14077 .......................................................................C........ 14156

CmRV1_2               13768 CTATTCTTTTCCTAACATTTTCCCTTCATTGTCACTTAGCTCCTCAATCCACACAAGAATTCAATCATTCAAACAAGATC 13847
Rhesus macaque chr14  14157 ................................................................................ 14236

CmRV1_2               13848 ATTTTATTATAATATAAATT 13867
Rhesus macaque chr14  14237 .................... 14256
```

Figure S14. Alignment of the nucleotide sequence between CmRV1_2 and its ortholog in rhesus macaque. The proviral sequences with the flanking sequences, 5’ and 3’ LTRs, *gag*, *pol*, and *env* genes are shown. Flanking 4-bp target site duplication (TSD) sequences are shown in gray. Dots indicate identical nucleotide sequences, and dashes indicate the missing nucleotides. N indicates nucleotide gaps or not determined. Nucleotide sequences encoding Env are highlighted in green.

CmRV-trunc1

```
  1 METPTLQGPAQDKTLSYLTPCLLLALLPCAAGSNNPRGHMKPPGYIFIDNNPHRPYNLTW  60
 61 QVINFNNHEVLGETSKIAPIGTWFPDLYFNLDKVAGVNEMEGGEWRKQARRVSISRNGFY 120
121 ACPGFRTGDMEKTCGDITHLYCYSWSCVTNNDGEWKWATKPWYITMSFVQPCTRTRYSKN 180
181 CNLVRIKFEDAAKSDNSWITGLIWGLYLYQKPLYGIPIQIRLLVDPDIAPVAVGLNQVLS 240
241 EEKKPPVPIPEKPQPKAPQSTSPPLISASSEYTSSAQNVTR 281
```

PNRC2-ChRV

```
  1 METSSLQGPPQDKTFPCLTPHLLLTLFLPWVVGSSNPHQPYRLTWQITNFKTHEVLNETS  60
 61 HVAPLNTWFPDLYFNLDKIAMIDEMEGGEWRKQARRVSLSRNGFYVCPGFRTGPMKKTCG 120
121 EIMSLYCASWSCVTTNDGEWKWKTQPWYLTMSYVQPCTRTRYSATCNLIRVKFEEAAKTD 180
181 PRWTTGLIWGLNLYQTPAAGLPIQIRLLVNPVSASVPVGPNPVLTGRAPSQ 231
```

PNRC2-BoRV

```
  1 METSSLQGPPQDKTFPCLTPHLLLTLFLPWVVGSSNPHQPYRLTWQITNFKTHEVLNETS  60
 61 HVAPLNTWFPDLYFNLDKIAMTDEMEGGEWRKQARRVSLSRNGFYVCPGFRTGPMKKTCG 120
121 EIMSLYCASWSCVTTNDGEWKWKTQPWYLTMSYVQPCTRTRYSATCNLIRVKFEEAAKTD 180
181 PRWTTGLIWGLNLYQTPAAGLPIQIRLLVNPVSASVPVGPNPVLTGRAPSQ 231
```

BoRV-7

```
  1 METSSLQGPPQDKTRPCLTPHLLLALFLPWVIGSSNPHQPYRLTWQITNFETHEVLNETS  60
 61 HVAPLNTWFPDLYFNLDKIAMIDEMEGGEWRKQARRVSLSRNGFYVCPGFRTGPMKKTCG 120
121 EIMSLYCASWSCVTTNDGEWKWETQPWYLTMSYVQPCTRTRYSATCNLIRVKFEEAAKTD 180
181 PRWTTGLIWGLNLYQTPASGLPIQIRLLVNPVSASVPVGPNPVLTGRAPS 230
```

GoRV-trunc1

```
  1 METSSLRGPPQDKTLPCLTPHLLLTLALFLPWVTGNGNPYQPYQLTWQVTNFETHEVLNE  60
 61 TSHVAPLNTWFPDLYFNLDQIARINEMEGGEWRKNARRVSLSRNGFYACPGFRTGPMKKT 120
121 CGEITSLYCASWSCVTTNDGESKWKTQPWYVTMTYVRPCTRTRYSATCNLIRIKFEEAAK 180
181 TDSRWTSGLIWGLNLYQTPAFGVPIQIKLIVNPASAPVPIGPNQVLTGKAPPSSRSRQKV 240
241 PTAAPPTSSPSALPGTTPLPPDPETSNRLFHLIRGAYLALNQTKPESTASCWLCLATGPP 300
301 YYEGIASVSNFTNSTNHSGCAWERHKKLTLSEVSGSGTCIGKVPPSHQHLCNVTLTVPST 360
361 SHYLIPSGTDWWACDTGLTPCLPLAYPQLSSTTVKIIVHCYRLCPEFITKLENLSSSSLS 420
421 KNISLEGRENLFPSPSLLC 439
```

CERV2 provirus (ChrY)

```
  1 METSSLQGPPQDKTLPCLTPHLLLALFLPWVIGSSNPHQPYRLTWQITNFETHEVLNETS  60
 61 HIAPLNTWFPDLYFNLDKIAMINEMEGGEWRKQARKVSLSRNGFYVCPGFRTGPMKKTCG 120
121 EIMSLYCASWSCVTTNDGEWKWKTQLWYVTMSYVQPCTRTRYSATCNLICVKFEEAAKTD 180
181 PRWTTGLI 188
```

Figure S15. DNA synthesis of ERV *Env* genes. All sequences represent truncated Env proteins from primates including crab-eating macaque (CmRV-trunc1), chimpanzee [PNRC2-ChRV, CERV2 provirus (ChrY)], bonobo (PNRC2-BoRV, BoRV-7), and gorilla (GoRV-trunc1). Chr, chromosome.

Table S1. **Characteristics of primate ERVs used in this study.**

ERV	Host	Chromosome: position	Length of amino acids	Properties	Accession no.
PNRC2-BoRV	Bonobo	chr1: 23161794-23162489	231	Truncated Env	NC_048239.1
BoRV-7	Bonobo	chr7: 59391503-59392195	230	Truncated Env	NC_048246.1
PNRC2-ChRV	Chimpanzee	chr1: 22537082-22537777	231	Truncated Env	XM_016938622.2
CERV2	Chimpanzee		651	Infectious Env (extinct)	
CERV2-trunc	Chimpanzee		188	Truncated Env (artificial)	
CERV2 provirus (ChrY)	Chimpanzee	chrY: 20227411-20227977	188	Truncated Env	AY692037.1
CmRV1	Crab-eating macaque		626	Intact full Env	KT330768.1 KT330663.1
CmRV1_1	Crab-eating macaque	chr13: 2014693-2016573	626	Intact full Env	NC_022284.1
CmRV1_2	Crab-eating macaque	chr14: 109894221-109896104	626	Intact full Env	NC_022285.1
CmRV-trunc1	Crab-eating macaque	chr1: 104267714-104268559	281	Truncated Env	XM_015429990.1
RmRV1	Rhesus macaque	chr11: 104718341-104720221	626	Intact full Env	NC_041764.1
RmRV1	Rhesus macaque		626	Intact full Env	KT330634.1
RmRV-trunc1	Rhesus macaque	chr1: 103221116-103221961	281	Truncated Env	NC_041754.1
RhERV2-A	Rhesus macaque		626	Infectious Env (extinct)	
GoRV-trunc1	Gorilla	chr1: 45572110-45573429	439	Truncated Env	NC_044602.1

PNRC2, proline rich nuclear receptor coactivator 2; trunc, truncated, ChRV, chimpanzee ERV; BoRV, bonobo ERV; CERV, chimpanzee ERV; CmRV, crab-eating macaque ERV; RmRV, rhesus macaque ERV, RhERV, rhesus macaque ERV; GoRV, gorilla ERV. CmRV *env* genes were found in two loci in the crab-eating macaque genome.

Table S2. Infection tropism of RmRV1/CmRV1, CERV2, and FeLV-D (ON-T) Env-pseudotyped viruses

Species	Cell	RmRV1/CmRV1	CERV2	FeLV-D (ON-T)
Human (*Homo sapiens*)	HepG2	(9.93 ± 2.34) x 10^{2}	(2.19 ± 0.71) x 10^{3}	(1.37 ± 0.22) x 10^{4}
	HEK293T	(4.02 ± 1.86) x 10^{3}	(9.67 ± 0.72) x 10^{2}	(3.65 ± 0.21) x 10^{4}
Cat (*Felis catus*)	CRFK	(2.56 ± 0.15) x 10^{3}	(1.24 ± 0.04) x 10^{3}	(6.75 ± 0.70) x 10^{2}
African green monkey (*Chlorocebus aethiops*)	Vero	(7.47 ± 1.64) x 10^{1}	(5.00 ± 0.20) x 10^{1}	(7.60 ± 2.40) x 10^{1}
	Cos7	(2.28 ± 0.51) x 10^{2}	(1.35 ± 0.24) x 10^{3}	(1.24 ± 0.20) x 10^{2}
Earth-colored Mouse (*Mus dunni*)	MDTF	0	6.40 ± 5.20	0
Golden hamster (*Mesocricetus auratus*)	BHK21	0	0	0
Guinea pig (*Cavia porcellus*)	104C1	(1.58 ± 0.26) x 10^{3}	(5.16 ± 0.40) x 10^{2}	0
Dog (*Canis familiaris*)	KwDM	(9.79 ± 0.45) x 10^{3}	(1.37 ± 0.02) x 10^{3}	(5.99 ± 0.03) x 10^{3}
Cattle (*Bos taurus*)	MDBK	(1.66 ± 0.20) x 10^{3}	(1.63 ± 0.00) x 10^{3}	(3.78 ± 0.03) x 10^{3}

Titers are expressed in infectious units/mL. Infective titers with standard deviations were means of three independent infection experiments.

Dataset S1. SRA Run accession, genome file, and gene annotation file used for the mapping analysis.

Dataset S2. Sequence accession numbers used for phylogenetic analyses.