

Title: Multiple invasions of an infectious retrovirus in cat genomes

Authors: Sayumi Shimode¹, So Nakagawa² & Takayuki Miyazawa^{1*}

Author affiliations

¹Laboratory of Signal Transduction, Department of Cell Biology, Institute for Virus Research,
Kyoto University, 53 Shogoin-Kawaharacho, Sakyo-ku, Kyoto 606-8501, Japan.

²Department of Molecular Life Science, Tokai University School of Medicine, 143 Shimokasuya,
Isehara, Kanagawa 259-1193, Japan.

Keywords: endogenous retrovirus, recombination, domestic cat

Numbers of Tables and Figures combined: 9

***Corresponding author:** Dr. Takayuki Miyazawa, Laboratory of Signal Transduction,
Department of Cell Biology, Institute for Virus Research, Kyoto University, 53
Shogoin-Kawaharacho, Sakyo-ku, Kyoto 606-8507, Japan.

Tel/Fax: +81-(0)75-751-4814

E-mail: takavet@gmail.com

Abstract

Endogenous retroviruses (ERVs) are remnants of ancient retroviral infections of host germ-line cells. While most ERVs are defective, some are active and express viral proteins. The RD-114 virus is a replication-competent feline ERV, and several feline cell lines produce infectious RD-114 viral particles. All domestic cats are considered to have an ERV locus encoding a replication-competent RD-114 virus in their genomes; however, the locus has not been identified. In this study, we investigated RD-114 virus-related proviral loci in genomes of domestic cats, and found that none were capable of producing infectious viruses. We also found that all domestic cats have an RD-114 virus-related sequence on chromosome C2, termed RDRS C2a, but populations of the other RDRSs are different depending on the regions where cats live or breed. Our results indicate that RDRS C2a, the oldest RD-114-related provirus, entered the host genome before an ancestor of domestic cats started diverging and the other new RDRSs might have integrated into migrating cats in Europe. We also show that infectious RD-114 virus can be resurrected by the recombination between two non-infectious RDRSs. From these data, we conclude that cats do not harbor infectious RD-114 viral loci in their genomes and RD-114-related viruses invaded cat genomes multiple times.

Endogenous retroviruses (ERVs) are retroviruses which have been integrated into the host genome of germ-line cells, comprising approximately 10 % of the host genome in mammals¹. ERVs behave as host genes and are transmitted from parents to offspring in a Mendelian manner. Although most ERVs are inactivated through the accumulation of mutations and deletions, leading to an introduction of stop codons within coding sequences, some ERVs are active and have the potential to produce infectious viral particles.

The RD-114 virus is a replication-competent feline ERV² and several feline cell lines produce infectious RD-114 viral particles³. The RD-114 virus was first isolated from human rhabdomyosarcoma (RD) cells during passages through fetal cat brains and was mistakenly regarded as the first human retrovirus⁴. Subsequently, it was revealed that cellular RNA from cats contain sequences homologous to RD-114 viral RNA^{5,6}, leading to the discovery of RD-114 virus-related proviral sequences in the cat genome^{7,8,9,10}. In addition, several groups reported that infectious RD-114 viral particles were produced, either spontaneously or after chemical induction, from cells originating from domestic cats^{11,12,13}. From these data, it was concluded that the RD-114 virus was an active ERV of cats, and not an exogenous human retrovirus.

Several research groups have searched for the active RD-114 proviral loci^{2,14,15}. Spodick *et al.* identified thirteen RD-114 virus-related clones, consisting of conserved *gag-pol* and unrelated *env* regions from six domestic cat genomes¹⁴. At the same time, another group identified nine endogenous RD-114 virus-related sequence clones representing at least eight distinct loci from cat DNA libraries². Although the *gag-pol* region of the clones was conserved, the *env* region displayed considerable divergence from the replication-competent RD-114 virus clone derived

from human RD cells, both in size and sequence homology, coinciding with previous results. In their followup study, feline chromosomes which contain RD-114 virus-related segments were identified with variation of distinct viral segments among individual cats¹⁵. These two groups identified the restriction endonuclease-digested genome fragments corresponding to fragments of the infectious RD-114 virus clone in cat genomic DNA and concluded that the active infectious RD-114 viral locus may be a single copy. The former group detected a 4.8-kb *Pst*I-digested *pol-env* fragment in all cats genomic DNAs examined¹⁴, and the latter group detected a 1.0-kb *Bam*HI-digested *gag* fragment^{2,15}. Although the latter fragment was found to be located on chromosome B3¹⁵, the locus of the former fragment was not identified. Further it has not been confirmed whether both of them were present on the same single locus as the complete full-length provirus. After these studies, a new endogenous type C retrovirus named *Felis catus* endogenous retrovirus (FcEV) which consists of the RD-114 virus-related *gag-pol* gene and unrelated *env* gene was discovered in all domestic cats genomes examined¹⁶. They reported that there were fifteen to twenty FcEVs in domestic cat genomes and suggested that all or most RD-114 virus-related sequences (RDRSs) possessing unrelated *env* genes are more likely to be FcEV integrations^{16,17}. Therefore, as mentioned above, the RD-114 proviral locus encoding an infectious RD-114 virus has not been identified.

In this study, we searched for loci encoding infectious RD-114 virus. Contrary to our expectations, we could not find any RDRSs capable of producing infectious RD-114 virus, but discovered that a RD-114 virus-related virus (RDRV) nearly identical to the RD-114 virus can be produced by recombination between two defective RD-114 virus-related proviruses. We also

noticed that most RDRSs have not been fixed in domestic cat genomes, suggesting multiple invasions of cat genomes by the ancient RDRV. Our study sheds light on host-virus interactions and provides novel genetic markers of domestic cats available for tracing footsteps of cats.

Results

Characterization of RDRS loci in the genomes of domestic cats. Firstly, we searched for RD-114 proviral loci in the genome database of domestic cats (*Felis_catus_6.2*) using the SSEARCH program¹⁸ with the complete genome sequence of RD-114 strain CRT1 (GenBank accession number: AB559882.1)¹⁹ as a query; however, we could not identify any proviruses which were identical to the sequences of three previously reported infectious RD-114 virus clones²⁰. Instead, we found one defective RDRS locus on feline chromosome C2 and designated it as RDRS C2a (proviruses were named after the chromosome numbers on which they are located). Quite recently, by using the sequences of RD-114 virus infectious clones as queries, Tamazian *et al.* reported that there are 12 RD-114 virus-like elements with more than 50 % sequence identity, which cover at least 50 % of query sequences in domestic cat genomes²¹. However, their sequences are not available for further analyses. Next, we investigated whether RDRS(s) capable of inducing RD-114 virus are present in all cat genomes by Southern blot analysis. By using *HindIII* and a *gag* probe, we detected a 2.6-kb fragment corresponding to the *HindIII* fragment of RD-114 virus molecular clones in RD-114 virus producer Crandell-Rees feline kidney (CRFK) cells and RD-114 virus non-producer cell lines such as G355-5 (a feline fetal astrocyte cell line), *Felis catus* whole fetus-4 (fcwf-4) (a feline macrophage-like cell line)

and FeT-J (an interleukin-2-independent feline T-cell line) cells (Fig. 1B). By using *HindIII* and
 an *env* probe, we detected a 3.2-kb fragment corresponding to the *HindIII* fragment of RD-114
 virus molecular clones only in CRFK cells but not RD-114 virus non-producer cell lines (Fig.
 1D). In accordance with the previous studies^{2,14,15}, there are polymorphisms present in RD-114
 virus-related sequences with a conserved *env* region. By southern blotting hybridization using
SacII, *EcoRI* or *SphI* and probes for *gag*, *pol* and *env* genes, we revealed that RD-114
 virus-related *env* copy numbers were less than *gag-pol* copy numbers (Fig. 1B, C, D). Inverse
 nested PCR targeting RD-114 virus *env* identified three proviruses (designated as RDRS A2, C2a,
 and C1, respectively) in CRFK cells (Fig. 2 and Fig. S1). We also identified three proviruses
 (RDRS C2a, E3 and D4) in primary feline peripheral blood mononuclear cells (PBMCs) and two
 proviruses (RDRS C2a and C2b) in FER cells (feline embryonic fibroblasts that produce
 RD-114-related virus) (Fig. 2 and Fig. S1). Again, these proviruses were not identical to the
 sequences of infectious RD-114 virus clones. Four of these proviruses (RDRS A2, C1, D4 and
 C2b) have the entire open reading frames (ORFs) for *env* and two of them (RDRS C1 and C2b)
 have intact ORFs for all viral genes (*gag*, *pol* and *env*) (Fig. 2). We found that the 3.2-kb *HindIII*
 fragment containing *env* could be generated by digestion of four RDRSs (RDRS A2, C1, D4 and
 C2b) which have the intact ORF for *env* gene. RDRS C2a of CRFK, G355-5 cells and feline
 PBMCs (*n*=2) have stop codons within all viral genes in the same fashion. Phylogenetic analysis
 of RDRSs and RD-114 virus clones indicated that RDRS C2a belongs to a separated group from
 the other RDRSs and infectious RD-114 virus clones (Fig. 3). In feline cell lines, although all
 lines have RDRS C2a, RD-114 viral producer cell lines do not have any other loci in common

(Table 1). These data suggest that a single RDRS is not responsible for the production of RD-114 viral particles. We further investigated cat populations for the possession of RDRSs in their genomes. Similar to the data obtained from cell lines, we also found that all cats have RDRS C2a in common, while the possession patterns of the other RDRS loci varied depending on the regions where cats live or breed (Fig. 4A).

Timelines for acquisitions of RDRSs in feline species. We next estimated the integration time of RDRSs. The divergence accumulated between 5' and 3' LTRs over time was previously used as a molecular clock to investigate the integration time of ERV²². Thus, we tried to estimate the integration time of the RDRSs by assessing the sequence differences between 5' and 3' LTRs. The proposed rate of 1.2×10^{-8} substitutions/site/year in cats was adopted for calculating proviral insertion time²³. Four of the six proviruses (RDRS A2, E3, D4 and C2b) have identical 5' and 3' LTRs, which are suggestive of relatively recent integrations into the host genome, less than 0.2 million years ago (MYA), possibly by RDRV infection (Fig. 2C). On the other hand, RDRS C1 and C2a have one and eight nucleotide differences, resulting in an estimated time of integration of RDRS C1 and C2a around 0.2 and 1.6 MYA, respectively (Fig. 2C). These periods are after domestic cats separated from the Sand cat lineage²⁴. The time of ERV integration can be also estimated by determining whether phylogenetically closely related animal species share these proviruses in the same genome location. The oldest RDRV provirus C2a was not present in the genome of the Leopard cat (*Prionailurus bengalensis*) ($n=4$) which inhabits Vietnam. Instead, we found a 54-bp sequence within the same locus. Intriguingly, the 54-bp sequence is part of a SINE element (Rebase ID: SINEC_Fc2), predicted by the Repeatmasker program

(<http://www.repeatmasker.org/>) (Fig. 5). The SINE-like sequence is defined by a tRNA-related region, harboring RNA polymerase III-specific internal promoter with its conserved regulatory elements, A-box and B-box, followed by a microsatellite region (TC)_n and by an A/T-rich tail with the polyadenylation signal AATAAA^{25,26}. The 54-bp sequence in the leopard cat genome contains the A-box, but the domestic cat's genome lacks this part of the SINE element. In other Felidae, including the Serval cat (*Leptailurus serval*), Snow leopard (*Panthera uncia*), and Tiger (*Panthera tigris*) (whole genome shotgun sequences, accession number: NW_006711997.1), the full length SINE elements were maintained in their genomes at the same location as the leopard cat, suggesting that SINEC_Fc2 integrated into the genome of the common ancestor of Felidae species and then RDRS C2a integration occurred, leading to the disruption of a part of the SINE sequence (Fig. 5). Anai *et al.* also reported the absence of the RD-114 viral *env* gene in the Tsushima cat (*Prionailurus bengalensis euptilurus*)^{27,28}. The domestic cat is estimated to have separated from the leopard cat lineage approximately 6.2 MYA²⁸. We hypothesize that the oldest RDRV C2a entered the ancestor of domestic cats from 6.2 to 0.2 MYA before the diversification of cat breeds.

Integration of RDRS coding for infectious-type *env*. Because RDRS sequences vary depending on where cats live or breed, it is possible that unidentified RDRS loci are present in domestic cat genomes of breeds investigated above. We developed a PCR method to differentiate infectious-type *env* on any loci from the defective C2a-type *env* (Fig. 4B). RDRS C2a has a 15-bp insertion in the *env* coding region, whereas the other RDRSs do not. We designed a reverse primer to step over the region of the 15-bp insertion present in RDRS C2a (Fig. 4B). As a result,

we revealed that 40 and 56 % of domestic cats in Europe and North America have infectious-type *env*, respectively, whereas almost all domestic cats in Asia do not (Fig. 4A). Asian domestic cats containing infectious-type *env* were mongrel cats and their origins were not identified. Interestingly, none of the Asian cats have any RDRSs harboring infectious-type *env* identified in this study. Similarly, most Scottish fold cats have RDRS harboring infectious-type *env*; however, they are unknown RDRSs. These data may indicate that unidentified RDRS(s) with infectious-type *env* are present in Scottish fold cats. Domestic cats (*Felis silvestris catus*) originate from the African wildcat (*Felis silvestris lybica*), a wild cat living in the Middle East, and then spread over the world to Europe or a different route to Asia²⁴ (Fig. 4C). Our results suggested that the newer RDRVs (*i.e.* RDRSs possessing infectious-type *env*) might have infected domestic cats after cats started moving from the Middle East. There is a possibility that some additional loci with the infectious-type *env* were acquired in cats through introgression of newer RDRSs into their genomes by crossbreeding between different feline species or cat breeds^{21,29,30,31}. Although the original cat breeds of CRFK, 3201 (feline thymic lymphoma) and FER cells possessing the newer RDRSs are unclear, these cell lines were established in the United States or the United Kingdom^{32,33,34}, suggesting that they originated from cats living in these countries.

Infectivity of RDRSs. Among the RDRS loci, RDRS C1 and C2b have intact ORFs for *gag*, *pol* and *env* genes, and RDRS C1 was most similar to the sequences of infectious RD-114 virus clones²⁰. RDRS C1 has five differences, when compared with an infectious RD-114 virus clone, termed pSc3c; LTR has one additional direct repeat (termed Δ DR-A), one point mutation at primer

binding site (the “a” to “g” change at position 491 of pSc3c), and three point mutations in *pol* gene (“a” to “g” at 3748, “t” to “c” at 5394 and “a” to “g” at 5719) resulting in amino acid substitutions (Fig. 6A). To examine if RDRS C1 encodes infectious RD-114-like virus, we transfected human embryonic kidney (HEK) 293T cells with a plasmid encoding RDRS C1. We then collected culture supernatants at 48 hours post transfection, and inoculated them onto HEK293T cells transduced with the LacZ marker gene [HEK293T(LacZ) cells]. Virus titers produced in the culture supernatants were measured by the LacZ marker rescue assay as described previously³⁵. As a result, we found that RDRS C1 was not infectious. We further evaluated the effects of the three point mutations in the *pol* region on the infectivity of RDRS C1. We constructed single mutants, which were made by site-directed mutagenesis³⁶. Two mutants were not infectious, whereas one mutant (5394t mutant) with a mutation introduced in the integrase core domain (Pfam accession number: PF00665), predicted by Pfam, (<http://pfam.sanger.ac.uk/>) had infectivity as an RD-114 virus clone (Fig. 6B and C). From these results, we conclude that the “t” at 5394 is crucial for infectivity for RDRS C1 infectivity. The same mutation in RDRS C1 has also been observed in the sequences of two unrelated cats.

In addition to CRFK cells, FER cells produce infectious RD-114-related viral particles³. Because FER cells have RDRS C2b which has intact ORFs for *gag*, *pol* and *env* genes (Fig. 2), we suspected that RDRS C2b may encode an infectious RD-114-like virus. RDRS C2b has a 9-bp insertion at the *pol* coding region, resulting in the insertion of three amino acids in comparison with pSc3c (Fig. S2). We examined the infectivity of RDRS C2b in the same way as described

above. Consistent with results of the other RDRSs identified in this study, we found that RDRS C2b was not infectious (Fig. S2).

Recombination of RDRSs. Because we could not find any loci which encode infectious RD-114 virus, we hypothesized that recombination between RDRSs resulted in the production of infectious RD-114 viral particles. We attempted to reproduce the recombination event by transfection of two RDRS plasmids. Because CRFK cells, from which infectious RD-114 virus is produced, have RDRS A2, C2a and C1 loci, we suspected that an infectious RDRV might be generated by recombination between RDRS A2 and C1. We transfected two plasmids encoding RDRS A2 and C1, and succeeded in recovering replication-competent viral particles (Fig. 7A). By comparing sequences of RDRS A2 and C1, and the recombinant RDRV obtained by transfection of RDRS A2 and C1 (hereinafter referred to as RDRV AC), it was suggested that three changes had occurred (two in the middle of *pol* and one in *pol-env* region) (Fig 7B). Of the two single nucleotide differences between RDRV AC and pSc3c at positions 3224 and 4357, only the “c” to “a” at 3224 results in an amino acid substitution. The amino acid difference was not present in the predicted functional domains of Pol. The length of LTR of RDRV AC was 27 bp shorter than that of the recombination origin RDRS C1 at the repeat region (Fig. 7C). The 27-bp sequence was repeated in the U3 region of LTR in pSc3c, RDRS A2 and C1, and termed direct repeat A1 (DR-A1). Altogether, infectious RD-114 virus-like particles can be generated by recombination between defective RDRS A2 and C1.

Discussion

The RD-114 virus was originally isolated from human RD cells after passages in fetal cat brains⁴. Because the parent cell line (RD cells) does not contain the RD-114 virus genome, there has been speculation about the origin of the RD-114 virus. In previous studies, restriction endonuclease map analysis and Southern blot hybridization of cat cellular DNA revealed that there are approximately twenty RDRSs in domestic cat genomes but most have a deleted-*env* region^{2,14,15}. The 4.8 kb *Pst*I-digested fragment corresponding to a complete *env* region was detected in all cat DNAs examined as a single copy in the previous study; however, they did not isolate a clone containing a complete RD-114 provirus¹⁴. We confirmed that the *Pst*I-digested *env* fragments derived from RDRS C2a is similar to infectious RD-114 virus clones, pSc3c and pCRT1, in size (4.7 kb) with high sequence homology (95.7%, data not shown). These data suggest that the single locus in all cat genomes which Spodick *et al.* identified was in fact RDRS C2a and not the complete infectious RD-114 provirus. Another group identified eight RDRSs (Ren 18c, 20a, 8a, 7c, 10a and 6c) in domestic cat genomes² and found that the 1.0-kb *Bam*HI *gag* fragment specific to the RD-114 virus clones was present in cat genomes as a single copy on chromosome B3; however, domestic cats are polymorphic with respect to the presence or absence of this RD-114 virus-related segment^{2,15}. Although it has been proposed that all domestic cats harbor infectious RD-114 provirus in their genomes, the proviral locus has not been identified even after extensive mining of feline genome databases.

In this study, we identified six RDRSs in domestic cats genomes. These proviruses are not identical to the sequences of infectious molecular clones of RD-114 virus reported previously²⁰. RDRS C1, whose sequence is most similar those of RD-114 infectious clones, could not produce

infectious viral particles. The defect was ascribed to a single nucleotide mutation resulting in an amino acid substitution in the integrase core domain which plays a key role in catalysis³⁷. The sequences of RDRS LTRs are rich in diversity, suggesting that each RDRS entered the host genome at different time points. Five of the RDRSs (excluding RDRS C2a) have identical 6-bp target site duplications (TSDs) (Table 2) and there were no differences between 5' and 3' LTRs in RDRSs A2, C2b and E3. These data suggest that LTR-LTR recombination did not occur and RDRS invasions into cat genomes occurred relatively recently. Except for RDRS C2a, RDRSs have diverse TSDs, indicating that multiplication of RDRSs occurred through viral re-infection but not genome rearrangements. The mechanism of multiple invasions of RDRVs is unknown at present. Generally, it is considered that integration of ERVs into host genomes confers resistance to the infection of exogenous counterparts³⁸. However, it is possible that the expression of newly integrated RDRS is tightly regulated in cats, especially in germ-line cells, to allow multiple integrations of RDRV in cats.

Remarkably, number of RDRS proviral copies vary considerably with type of cell lines and cat breeds. All cats have RDRS C2a in common, but all proviral genes (*gag*, *pol* and *env*) are disrupted in RDRS C2a. The other RDRSs have not yet been fixed in domestic cat genomes. Furthermore, we found that most Asian cats are free from new RDRSs harboring infectious-type *env*.

Because domestic cats originated from the African wildcat living in the Middle East (*Felis silvestris lybica*)²⁴, we propose the following scenario (Fig 4C) for RDRV integration into domestic cat genomes in Europe and Asia. (i) Firstly the ancestral virus of RDRS C2a (RDRV

C2a) infected the domestic cat progenitor. (ii) The African wildcats were domesticated in the Middle East where agriculture started around 10,000 years ago. (iii) Then, some domesticated cats moved to Europe and America as ship's cats with Vikings or traders, and became common throughout Europe³⁹. Ship's cats could travel to different harbors for chance encounters with wildcats for interbreeding³⁹, diversifying the genomes of their offspring but at the same time putting themselves at risk for new RDRV infection and subsequent integration by infectious-type RDRSs. (iv) The other cats went to Asia as guardians of Buddhist shariha along well-established trade routes, termed the Silk Road³⁹. With no native wildcats that lived near this road to interbreed with, Oriental domestic cats soon began evolving in their own way³⁹ without the risk of infection by new RDRVs.

British Shorthair (BSH) was established in England from European Shorthair which first appeared in Sweden, and later American Shorthair (ASH) and American Curl (a variant of ASH) originated from BSH⁴⁰. Some of these cat breeds have RDRS E3 but others do not. These data indicate that RDRS E3 would be a useful genetic marker to trace the footsteps of these breeds from Europe to America. In future studies, we will investigate the presence of RDRSs in domestic cats living in various regions as well as closely related wild cats to further map cat migration.

Several feline cell lines including CRFK and FER cells produce RD-114 or RD-114-related viruses, but there are no proviruses identical to previously reported RD-114 viral sequences. Several novel retroviruses have been reported to be generated by the recombination of ERVs^{41,42,43}. Therefore, we hypothesized that RD-114 virus-like particles were generated by the

recombination between two defective RDRS proviruses. By co-transfection of two defective RDRS A2 and C1, we demonstrated that infectious RD-114-like viral particles were generated through recombination. The resultant recombinant, RDRV AC, and infectious clones of RD-114 virus were nearly identical in coding sequences; however, the LTR of RDRV AC was not identical to those of the infectious clones, pCRT1 and pSc3c. The changes in LTR length frequently occurred in repeat sequences during cell passages *in vitro* as observed in other gammaretroviruses such as porcine endogenous retrovirus^{44,45}. FER cells also produce RD-114-related viral particles³, despite the lack of RDRS A2 and C1 in their genomes (Table 1). Similar to the case of CRFK cells, we speculated that an RD-114-related virus could have been generated by recombination between RDRS C2b which has the infectious-type *env* and some other unidentified RDRSs or FcEV.

Retroviral recombination events occur by template switching during reverse transcription^{46,47}. Xenotropic murine leukemia virus-related virus (XMRV) is one such virus which has arisen through the recombination of two defective murine ERVs during passages of human prostate cancer xenografts in immunocompromised nude mice⁴⁸. The defective viruses recombined in human cells but not in mouse cells because the XMRV is xenotropic. Similarly, the first RD-114 virus may have arisen by recombination in human RD xenografts in cat brains^{4,8}. RD-114-related viruses have also been induced by chemicals in feline cells without xenotransplantation^{11,12,13}, where a RD-114-related virus may have arisen in feline cells without contact with cells from other species. In these cases, defective viral particles co-packaged with RDRS genome infected feline cells and recombined during reverse transcription. Resultant replication-competent viruses may

have expanded in the feline cells, because feline cells are permissive to RD-114 virus^{3,49}. Additionally, we should consider a possibility that RDRS C1 produces infectious viral particles by spontaneous mutation like Emv2 in C57BL mice^{50,51,52}. Emv2 has a defect in the *pol* gene, compromising the ability of the Emv2 to produce infectious virus particles⁵³. However, Emv2 can produce an infectious endogenous ecotropic murine leukemia virus (E-MLV) through the acquisition of a spontaneous mutation in the *pol* gene *in vivo* in aging C57BL mice⁵⁰ or by backcrossing between C57BL and E-MLV-negative mice⁵¹. Therefore, we cannot exclude the possibility that RDRS C1 produces infectious viral particles through spontaneous mutations *in vivo*.

In the previous study, we revealed ubiquitous expression of the RD-114 viral receptor, termed ASCT2, in domestic cat tissues⁴⁹, suggesting that RD-114 viral recombination events may occur in cats, and the infectious RD-114 virus may be generated in some groups of cats having RDRSs harboring infectious-type *env*. Recently, it was reported that in immune-compromised mice, ERVs in the host genome were resurrected and induced lymphoma^{54,55}. In these cases, receptors for infectious ERVs are expressed in tissues of mice and infectious ERVs re-infected the host cells and induced lymphoma possibly via insertion of the ERVs in the vicinity of cellular proto-oncogenes^{54,55}. In this study, most RDRSs are not yet fixed in cat genomes, but some of them have the potential to generate a recombinant virus which could re-infect the host cells. This recombinant virus may re-infect cat cells which do not express RDRS Env, leading to new integrations. Some of the new integrations may induce diseases in cats, if the integration occurred in the vicinity of proto-oncogenes. While it is still unknown whether RD-114 virus causes any

diseases in cats, our report that RD-114 can be generated by recombination will be useful for further investigation of the pathogenicity of this virus.

Methods

Ethics statement. All animal care and procedures that were performed in this study conformed to the guidelines for animal experiments at Institute for Virus Research in Kyoto University (IVRKU), and all experimental protocols were approved by the Committee on the Ethics of Animal Experiments of IVRKU. Blood samples of domestic cats were collected for diagnostic purpose in veterinary clinics. Genomic DNAs of Leopard cats, serval and snow leopard were stored samples which had been used in our previous studies^{56,57}.

Blood and tissue samples. Domestic cat blood samples for isolation of PBMCs were provided from Fujimura Animal Hospital (Osaka, Japan), Kyoritsu Seiyaku Corporation (Tokyo, Japan) and Veterinary Clinics in Japan. The Snow leopard and Serval cat tissue samples were provided from Sapporo Maruyama Zoo (Hokkaido, Japan). These animals died from feline parvovirus infection⁵⁷.

Cell cultures. FeT-J and 3201 cells were cultured in RPMI 1640 medium (Sigma-Aldrich, Tokyo, Japan) supplemented with 10 % heat-inactivated fetal calf serum (FCS), penicillin (100 IU/ml) and streptomycin (100 µg/ml) (Invitrogen, Carlsbad, CA). CRFK, G355-5, fcwf-4, FER, AH927 (feline embryonic fibroblast), HEK293T and TE671 (human rhabdomyosarcoma) cells

were cultured in Dulbecco's modified Eagle's medium (DMEM) (Sigma-Aldrich) supplemented with 10 % FCS, penicillin (100 IU/ml), and streptomycin (100 µg/ml) (Invitrogen).

Genomic DNA isolation and genomic PCR. Genomic DNA was extracted from cell lines and PBMCs using a QIAamp DNA Blood Mini Kit (QIAGEN, Valencia, CA). Genomic PCR was performed by using PrimeSTAR GXL polymerase (TaKaRa, Shiga, Japan) or Ex Taq polymerase (TaKaRa). The former was used for cloning RDRSs and the recombinant RDRV, and the later was used for screening the presence RDRS. Amplicons were analyzed by electrophoresis in 0.8 % or 2 % agarose gels. Primers used in this study are listed in Table S1.

Southern blotting. A total of 5 µg (or 3 µg) genomic DNA was isolated from PBMCs or cell lines, digested with *Hind*III for detection of *gag* and *env* genes, *Sac*II for detection of *gag* and *pol* genes, *Eco*RI for detection of *env* genes, or *Sph*I for detection of *gag* and *env* genes. Digested fragments were separated by a 1.0 % agarose gel and transferred to positive charged-nylon membrane (Roche Diagnostics, Indianapolis, IN). Hybridization was performed using digoxigenin (DIG)-labeled PCR probes and DIG easy Hyb (Roche) at appropriate temperatures (50 °C for *gag* and *pol*, 65 °C for *env* detection) for 16 hours. After hybridization, the membrane was washed with a low stringency buffer (0.1 % SDS and 2 × SSC) at room temperature followed by a high stringency buffer (0.1 % SDS and 0.1 × SSC) at 68 °C. The hybridized bands were visualized using CDP-Star reagent (Roche) following the manufacturer's instructions. Probes were prepared from an infectious molecular clone of RD-114 virus, termed pCRT1¹⁹, using

DIG-labeled PCR probe synthesis kit (Roche) according to the manufacturer's instructions with primers used listed in Table S1.

Inverse PCR. Genomic DNA was digested with *EcoRI* (New England Biolabs) or *SphI* (New England Biolabs) and 100 ng of digested genomes were ligated by T4 ligase (TOYOBO, Osaka, Japan) at 16 °C overnight. Ligated DNA was amplified with specific primers listed in Table S1 using PrimeSTAR GXL polymerase (TaKaRa). Second PCR products were cloned into a pCR4 Blunt TOPO vector (Invitrogen) and sequenced. Loci of RDRSs were determined by comparing flanking sequences with the genome database of domestic cats (*Felis_catus_6.2*).

Cloning of RDRSs. Full-length RDRSs were amplified using specific primers (listed in Table S1) on flanking regions and amplicons were cloned into pCR BluntII TOPO vector (Invitrogen).

Infection assay. HEK293T cells were transfected with 4 µg of RD-114 virus molecular clones or plasmids containing RDRSs using Lipofectamine2000 (Invitrogen). The pcDNA3.1 vector was used for mock transfection. Infectivities of RDRSs were examined by the LacZ marker rescue assay²⁸. Culture supernatants were collected 48 hours post-transfection and filtrated through 0.45 µm-pore-size filters and then inoculated onto HEK293T(LacZ) cells with 8 µg/ml polybrene. After inoculation, supernatants were collected every four days. TE671 cells were seeded in 96-well plates at 1×10^4 cells per well one day before infection and diluted supernatants of inoculated HEK293T(LacZ) cells were inoculated onto TE671 cells in the presence of 8 µg/ml

polybrene. Two days after inoculation, the inoculated cells were stained with X-Gal, and virus titers, expressed as focus forming units (f.f.u.)/ml, were determined as described previously^{35,47}. Assays were conducted in triplicate for each individual sample.

Preparation of RDRS C1 mutants. RDRS C1 mutants were made by *DpnI* method³⁶. PCR was performed using PrimeSTAR GXL Polymerase (TaKaRa) with primers listed in Table S1. RDRS C1 plasmid DNA was used as template. Amplicons were digested with *DpnI* to eliminate the parental plasmid DNA. Digested DNA fragments were purified and transformed into *Escherichia coli* DH5 α .

RDRV AC isolation. HEK293T cells were transfected with 4 μ g of pCRT1 or plasmids containing RDRSs using Lipofectamine2000 (Invitrogen). pcDNA3.1 vector (Invitrogen) was used for mock transfection. Culture supernatants were collected 48 hours post-transfection and filtrated through 0.45 μ m-pore-size filters and then inoculated into HEK293T cells with 8 μ g/ml polybrene (Sigma-Aldrich). After inoculation, supernatants were collected and filtrated every four days. TE671 cells were seeded in 6-well plates at 1×10^6 cells per well one day before infection and diluted supernatants of the inoculated HEK293T cells (36 d.p.i.) were inoculated onto TE671 cells in the presence of 8 μ g/ml polybrene. Six hours after inoculation, genomic DNAs were extracted and proviruses were amplified by PCR. PCR was performed with PrimeSTAR GXL polymerase (TaKaRa) and primers listed in Table S1, according to the manufacturer's instructions. The forward primer was designed in the R region of 5'LTR and

reverse primer was designed in the U5 region of 3'LTR (positions of primers were indicated in Fig. 7).

Sequence Analyses. Sequencing was performed by a commercial DNA sequencing service (FASMAC, Kanagawa, Japan). Alignment of nucleotide sequences was computed using L-INS-i program in MAFFT version 7^{58,59}, and their phylogeny was inferred by RAxML version 7.2.6 that is based on the maximum likelihood method⁵⁷.

GenBank accession number. RDRS A2, C2a, C1, E3, D4 and C2b are deposited in GenBank as LC005744- LC005749.

REFERENCES

1. Gifford, R. & Tristem, M. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* **26**, 291-315 (2003).
2. Reeves, R. H. & O'Brien, S. J. Molecular genetic characterization of the RD-114 gene family of endogenous feline retroviral sequences. *J Virol* **52**, 164-171 (1984).
3. Okada, M., Yoshikawa, R., Shojima, T., Baba, K. & Miyazawa, T. Susceptibility and production of a feline endogenous retrovirus (RD-114 virus) in various feline cell lines. *Virus Res* **155**, 268-273 (2011).
4. McAllister, R. M. *et al.* C-type virus released from cultured human rhabdomyosarcoma cells. *Nat New Biol* **235**, 3-6 (1972).

- 433 5. Okabe, H., Gilden, R. V. & Hatanaka, M. Extensive homology of RD114 virus DNA with
434 RNA of feline cell origin. *Nat New Biol* **244**, 54-56 (1973).
- 435 6. Okabe, H., Gilden, R.V. & Hatanaka, M. RD 114 virus-specific sequences in feline cellular
436 RNA: detection and characterization. *J Virol* **12**, 984-994 (1973).
- 437 7. Baluda, M. A. & Roy-Burman, P. Partial characterization of RD114 virus by DNA-RNA
438 hybridization studies. *Nat New Biol* **244**, 59-62 (1973).
- 439 8. Neiman, P. E. Measurement of RD114 virus nucleotide sequences in feline cellular DNA.
440 *Nat New Biol* **244**, 62-64 (1973).
- 441 9. Gillespie, D., Gillespie, S., Gallo, R. C., East, J. L. & Dmochowski, L. Genetic origin of
442 RD114 and other RNA tumour viruses assayed by molecular hybridization. *Nat New Biol* **244**,
443 51-54 (1973).
- 444 10. Benveniste, R. E. & Todaro, G. J. Segregation of RD-114 AND FeL-V-related sequences in
445 crosses between domestic cat and leopard cat. *Nature* **257**, 506-508 (1975).
- 446 11. Livingston, D. M. & Todaro, G. J. Endogenous type C virus from a cat cell clone with
447 properties distinct from previously described feline type C virus. *Virology* **53**, 142-151 (1973).
- 448 12. Fischinger, P. J., Peebles, P. T., Nomura, S. & Haapala, D. K. Isolation of RD-114-like
449 oncornavirus from a cat cell line. *J Virol* **11**, 978-985 (1973).
- 450 13. Todaro, G. J., Benveniste, R. E., Lieber, M. M. & Livingston, D. M. Infectious type C viruses
451 released by normal cat embryo cells. *Virology* **55**, 506-515 (1973).
- 452 14. Spodick, D. A., Soe, L. H. & Roy-Burman, P. Genetic analysis of the feline RD-114
453 retrovirus-related endogenous elements. *Virus Res* **1**, 543-555 (1984).

454 15. Reeves, R. H., Nash, W. G. & O'Brien, S. J. Genetic mapping of endogenous RD-114
455 retroviral sequences of domestic cats. *J Virol* **56**, 303-306 (1985).

456 16. van der Kuyl, A. C., Dekker, J. T. & Goudsmit, J. Discovery of a new endogenous type C
457 retrovirus (FcEV) in cats: Evidence for RD-114 being an FcEV^{Gag-Pol}/baboon endogenous virus
458 BaEV^{Env} recombinant. *J Virol* **73**, 7994-8002 (1999).

459 17. Kim, F. J., Battini, J. L., Manel, N. & Sitbon, M. Emergence of vertebrate retroviruses and
460 envelope capture. *Virology* **318**, 183-191 (2004).

461 18. Pearson, W. R. Searching protein sequence libraries: comparison of the sensitivity and
462 selectivity of the Smith-Waterman and FASTA algorithms. *Genomics* **11**, 635-650 (1991).

463 19. Yoshikawa, R., Sato, E., Igarashi, T. & Miyazawa, T. Characterization of RD-114 virus
464 isolated from a commercial canine vaccine manufactured using CRFK cells. *J Clin Microbiol* **48**,
465 3366-3369 (2010).

466 20. Shimode, S. *et al.* Sequence comparison of three infectious molecular clones of RD-114
467 virus. *Virus Genes* **45**, 393-397 (2012).

468 21. Tamazian, G. *et al.* Annotated features of domestic cat - *Felis catus* genome. *Gigascience* **3**,
469 13 (2014).

470 22. Martins, H. & Villesen, P. Improved integration time estimation of endogenous retroviruses
471 with phylogenetic data. *PLoS One* **6**, e14745 (2011).

472 23. Lopez, J. V., Culver, M., Stephens, J. C., Johnson, W. E. & O'Brien, S. J. Rates of nuclear and
473 cytoplasmic mitochondrial DNA sequence divergence in mammals. *Mol Biol Evol* **14**, 277-286
474 (1997).

- 475 24. Driscoll, C. A. *et al.* The Near Eastern origin of cat domestication. *Science* **317**, 519-523
476 (2007).
- 477 25. Okada, N. SINEs: Short interspersed repeated elements of the eukaryotic genome. *Trends*
478 *Ecol Evol* **6**, 358-361 (1991)
- 479 26. Okada, N., Hamada, M., Ogiwara, I. & Ohshima, K. SINEs and LINEs share common 3'
480 sequences: a review. *Gene* **205**, 229-243 (1997).
- 481 27. Anai, Y. *et al.* Infectious endogenous retroviruses in cats and emergence of recombinant
482 viruses. *J Virol* **86**, 8634-8644 (2012).
- 483 28. Johnson, W. E *et al.* The late Miocene radiation of modern Felidae: a genetic assessment.
484 *Science* **311**, 73-77 (2006).
- 485 29. Beaumont, M. *et al.* Genetic diversity and introgression in the Scottish wildcat. *Mol Ecol* **10**,
486 319-336 (2001).
- 487 30. Randi, E., Pierpaoloni, M., Beaumont, M., Ragni, B. and Sforzi, A. Genetic identification of
488 wild and domestic cats (*Felis silvestris*) and their hybrids using Bayesian clustering methods. *Mol*
489 *Biol Evol* **18**, 1679-1693 (2001).
- 490 31. Oliveira, R., Godinho, R., Randi, E., Ferrand, N. and Alves, P. C. Molecular analysis of
491 hybridisation between wild and domestic cats (*Felis silvestris*) in Portugal: implications for
492 conservation. *Conservation Genetics* **9**, 1-11 (2008).
- 493 32. Crandell, R. A., Fabricant, C. G. & Nelson-Ress, W. A. Development, characterization, and
494 viral susceptibility of a feline (*Felis catus*) renal cell line (CRFK). *In Vitro* **9**, 176-185 (1973).

495 33. Rojko, J. L. *et al.* Feline lymphomas: immunological and cytochemical characterization.
496 *Nature* **19**, 656-658 (1978).

497 34. Jarrett, O., Laid, H. M. & Hay, D. Determinants of the host range of feline leukemia viruses.
498 *J Gen Virol* **20**, 169-175 (1973).

499 35. Sakaguchi, S., Okada, M., Shojima, T., Baba, K. & Miyazawa, T. Establishment of a LacZ
500 marker rescue assay to detect infectious RD114 virus. *J Vet Med Sci* **70**, 785-790 (2008).

501 36. Li, S. & Wilkinson, M. F. Site-directed mutagenesis: a two-step method using PCR and *DpnI*.
502 *Biotechniques* **23**, 588-590 (1997).

503 37. Craigie, R. HIV integrase, a brief overview from chemistry to therapeutics. *J Biol Chem* **276**,
504 23213-23216 (2001).

505 38. Palmarini, M., Mura, M. and Spencer, T. E. Endogenous betaretroviruses of sheep: teaching
506 new lessons in retroviral interference and adaptation. *J Gen Virol* **85**, 1-13 (2004).

507 39. Driscoll, C. A., Clutton-Brock, J., Kitchener, A. C. & O'Brien, S. J. The Taming of the cat.
508 Genetic and archaeological findings hint that wildcats became housecats earlier--and in a
509 different place--than previously thought. *Sci Am* **300**, 68-75 (2009).

510 40. Menotti-Raymond, M. *et al.* Patterns of molecular genetic variation among cat breeds.
511 *Genomics* **91**, 1-11 (2008).

512 41. Henzy, J. E., Gifford, R. J., Johnson, W. E. & Coffin, J. M. A novel recombinant retrovirus in
513 the genomes of modern birds combines features of avian and mammalian retroviruses. *J Virol* **88**,
514 2398-2405 (2014).

515 42. Delviks-Frankenberry, K. *et al.* Generation of multiple replication-competent retroviruses
516 through recombination between PreXMRV-1 and PreXMRV-2. *J Virol* **87**, 11525-11537 (2013).

517 43. Paprotka, T. *et al.* Recombinant origin of the retrovirus XMRV. *Science* **333**, 97-101 (2011).

518 44. Denner, J., Specke, V., Thiesen, U., Karlas, A. & Kurth, R. Genetic alterations of the long
519 terminal repeat of an ecotropic porcine endogenous retrovirus during passage in human cells.
520 *Virology* **314**, 125-133 (2003).

521 45. Denner, J. Recombinant porcine endogenous retroviruses (PERV-A/C): a new risk for
522 xenotransplantation? *Arch Virol* **153**, 1421-1426 (2008).

523 46. Delviks-Frankenberry, K. A., Nikolenko, G. N. & Pathak, V. K. The "Connection" Between
524 HIV Drug Resistance and RNase H. *Viruses* **2**, 1476-1503 (2010).

525 47. Takeuchi, Y. *et al.* Host range and interference studies of three classes of pig endogenous
526 retrovirus. *J Virol* **72**, 9986-9991 (1998).

527 48. Arias, M. & Fan, H. The saga of XMRV: a virus that infects human cells but is not a human
528 virus. *Emerg Microbes Infect* **3**, e25 (2014).

529 49. Shimode, S., Nakaoka, R., Shogen, H. & Miyazawa, T. Characterization of feline ASCT1
530 and ASCT2 as RD-114 virus receptor. *J Gen Virol* **94**, 1608-1612 (2013).

531 50. Odaka, T. Genetic transmission of endogenous N- and B-tropic murine leukemia viruses in
532 low-leukemic strain C57BL/6. *J Virol* **15**, 332-337 (1975).

533 51. Kozak, C. A. & Rowe, W. P. Genetic mapping of ecotropic murine leukemia virus-inducing
534 loci in six inbred strains. *J Exp Med* **155**, 524-534 (1982).

52. McCubrey, J. & Risser, R. Genetic interactions in the spontaneous production of endogenous murine leukemia virus in low leukemic mouse strains. *J Exp Med* **156**, 337-349 (1982).
53. King, S. R., Berson, B. J. & Risser, R. Mechanism of interaction between endogenous ecotropic murine leukemia viruses in (BALB/c X C57BL/6) hybrid cells. *Virology* **162**, 1-11 (1988).
54. Young, G. R. *et al.* Resurrection of endogenous retroviruses in antibody-deficient mice. *Nature* **491**, 774-778 (2012).
55. Yu, P. *et al.* Nucleic acid-sensing Toll-like receptors are essential for the control of endogenous retrovirus viremia and ERV-induced tumors. *Immunity* **37**, 867-879 (2012).
56. Nakamura, K. *et al.* Comparison of prevalence of feline herpesvirus type 1, calicivirus and parvovirus infections in domestic and leopard cats in Vietnam. *J Vet Med Sci* **61**, 1313-1315 (1999).
57. Sassa, Y. *et al.* Successive deaths of a captive snow leopard (*Uncia uncia*) and a serval (*Leptailurus serval*) by infection with feline panleukopenia virus at Sapporo Maruyama Zoo. *J Vet Med Sci* **73**, 491-494 (2011).
58. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780. (2013).
59. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690 (2006).

Acknowledgements

We are grateful to Dr. Masato Fujimura (Fujimura Animal Hospital) and Dr. Kazuo Kawakami (Kyoritsu Seiyaku Corporation) for providing domestic cat blood samples, Dr. Yasuhiro Takeuchi (University College London, London, U.K.) for providing TELCeB6/pRDF cells and TE671 cells chronically infected with RD-114 virus, and Dr. Kenji Baba (Yamaguchi University, Yamaguchi, Japan) for helpful advice. We thank Peter Gee (Kyoto University, Kyoto, Japan) for his generous help in the preparation of the manuscript. This work was supported by JSPS KAKENHI Grant Numbers 265589 and 25891023 (to SS and SN, respectively).

Author contributions

SS and TM designed the experiments. SS performed the experiments. SS, SN and TM analyzed data and wrote the manuscript. All authors read and approved the final manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.

Figure legends

Figure 1 Detection of RD-114 proviruses in the genomes of feline cell lines by Southern blotting analyses. (A) Restriction endonuclease maps of an infectious molecular clone of RD-114 virus, termed pSc3c. Probes for *gag*, *pol* and *env* genes are shown as bars. (B-D) Southern blotting analyses of several feline cell lines. Genomic DNAs were digested with *HindIII*, *SacII*, *EcoRI* or *SphI* and then subjected to Southern blotting analyses using *gag* (B), *pol* (C) and *env* (D) probes. *HindIII*-digested fragments corresponding to pSc3c are shown by white arrowheads. Asterisks indicate RD-114 viral *env* locus detected in common in all cell lines examined.

Figure 2 Structures of RDRSs. (A) Structures of the genomes of full length RDRSs. Open triangle indicates a deletion of nucleotides and filled triangle indicates an insertion of nucleotides. Green, orange and blue boxes indicate *gag*, *pol* and *env* ORFs, respectively. Numbers (%) above the diagrams indicate homologies with RD-114 infectious clones in each gene. (B) Schematic LTR structures of RDRSs and RD-114 virus clones, pSc3c and pCRT1. DR-A, direct repeat A; DR-B, direct repeat B; CAAT, CAAT box; TATA, TATA box; TSS, transcription start site; poly A, poly A signal. (C) Features of LTRs of RDRS proviruses and estimated integration time.

Figure 3 The maximum likelihood phylogenetic trees of RDRSs. We used nucleotide sequences of (A) *gag*, (B) *pol*, (C) *env* and (D) full length excluding LTR. The general time-reversible (GTR) model of nucleotide substitution with the addition of invariant sites (I) and

a gamma distribution of rates across sites (Γ) was used to infer the phylogenies. A2, RDRS A2; C2a, RDRS C2a; E3, RDRS E3; D4, RDRS D4; C2b, RDRS C2b; BaEV M7, Baboon endogenous virus strain M7.

Figure 4 Possession of RD-114 viral infectious-type *env* and RDRSs. (A) Population of RDRS proviruses in feline PBMC. Regions that the cats originally lived in were indicated based on a previous study⁴⁰. (B) Position of reverse primer to detect infectious-type *env* indicated by blue characters and arrowheads. Red characters are the RDRS C2a-specific insertion. Numbers are defined as positions of pSc3c. (C) Cat's journey and process of RDRV integration. Virus particle with blue hexagonal shape and orange hexagonal indicate the oldest RDRV and the new RDRVs, respectively. The picture of cats and the map were drawn by SS using the Microsoft PowerPoint.

Figure 5 Comparison of sequences of RDRS C2a integration sites in domestic cat's (*Felis catus*) and other Felidae species' genomes. (A) Alignment of sequences of 5' and 3' flanking regions of RDRS C2a in domestic cat (*Felis catus*) and their corresponding sites' sequences in leopard cat (*P. bengalensis*), Serval cat (*L. serval*), Snow leopard (*P. uncia*) and Tiger (*P. tigris*). Asterisks indicate conserved nucleotides between four species. Target site duplications (TSDs) for SINEC_Fc2-like sequence are shaded in black. (B) Schematic view of RDRS C2a integration site and adjacent sequences of Felidae genomes. Lime green arrow indicates RDRS C2a. Pink and pink-red arrows indicate SINEC_Fc2 and C SINEC_Fc2-like sequence, respectively, and

aqua arrow indicates LINE-like sequences. Short orange lines at both sides indicate primer positions used for PCR.

Figure 6 Infectivity of RDRS C1. (A) Differences between RDRS C1 and an infectious molecular clone, pSc3c. Numbers indicate nucleotide locations at pSc3c. Shaded regions are functional domains of *pol* region predicted by Pfam (PR, aspartyl protease; RT, reverse transcriptase; RH, RNaseH; IN, integrase core domain [Pfam accession number: PF00077, PF00078, PF00075 and PF00665, respectively]). Differences between RDRS C1 and pSc3c are indicated by vertical lines (point mutations) and filled triangle (insertion). RDRS C1 mutants have mutations marked by crosses. (B, C) LacZ marker rescue assay performed using TE671 cells as target cells. Infection with LacZ pseudotype viruses (RD-114 virus, RDRS C1 wild-type and RDRS C1 mutants) were visualized by X-Gal staining (B) and the virus titers were expressed as f.f.u./ml (C). Assays were performed in triplicate and the data are shown as the mean viral titers $\pm$ standard errors.

Figure 7 Recombination between RDRS A2 and C1. (A) Growth of RDRV AC in HEK293T cells. The RDRS plasmids were transfected into HEK293T cells and supernatants were inoculated into HEK293T(LacZ) cells, and then virus production was monitored by the LacZ marker rescue assay. Assays were performed in triplicate and the data are shown as the mean viral titers $\pm$ standard errors. (B) Comparison of nucleotide sequences of pSc3c, RDRS A2, C1 and RDRV AC virus. The nucleotide positions of pSc3c are shown. Dark grey arrowheads indicate

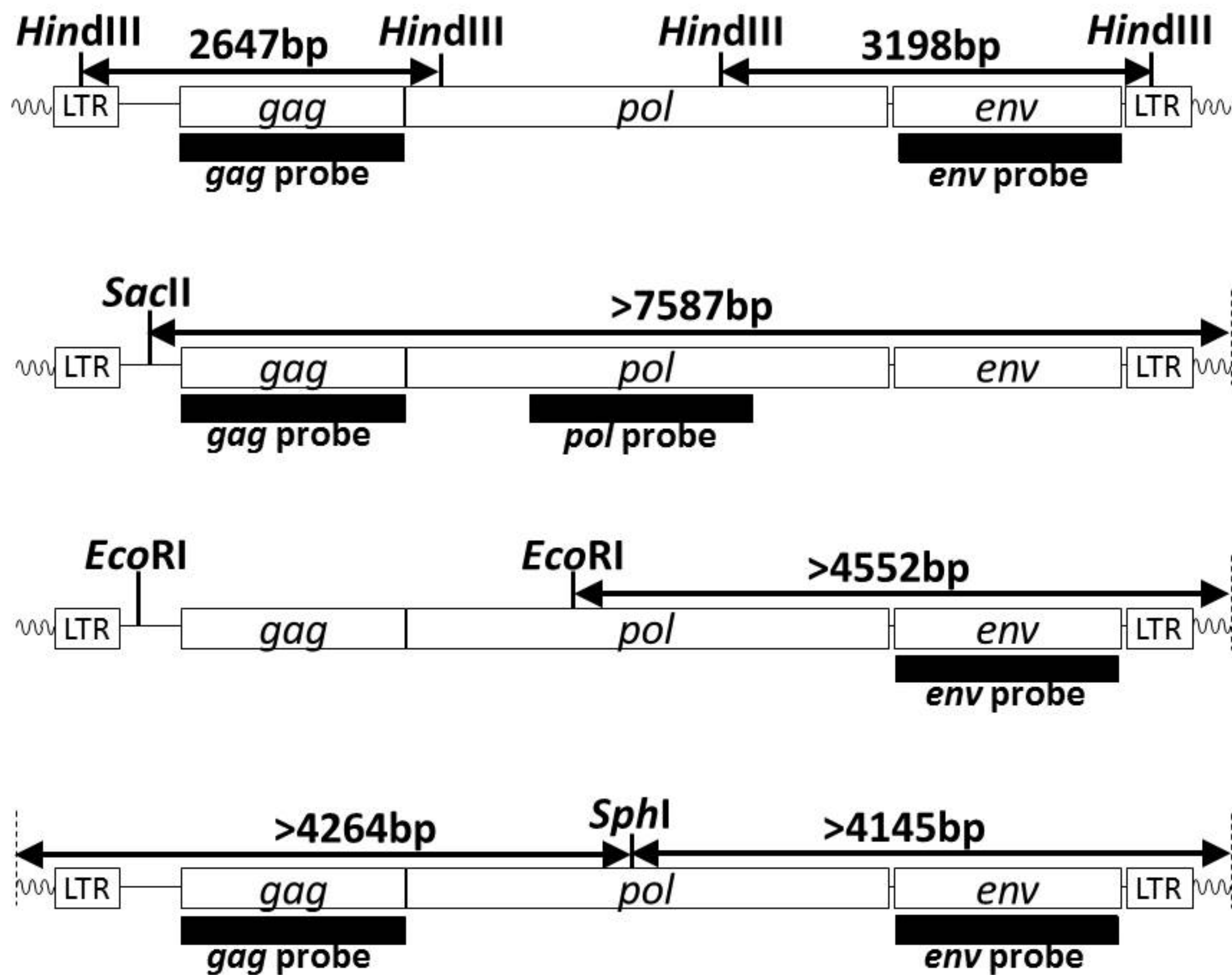
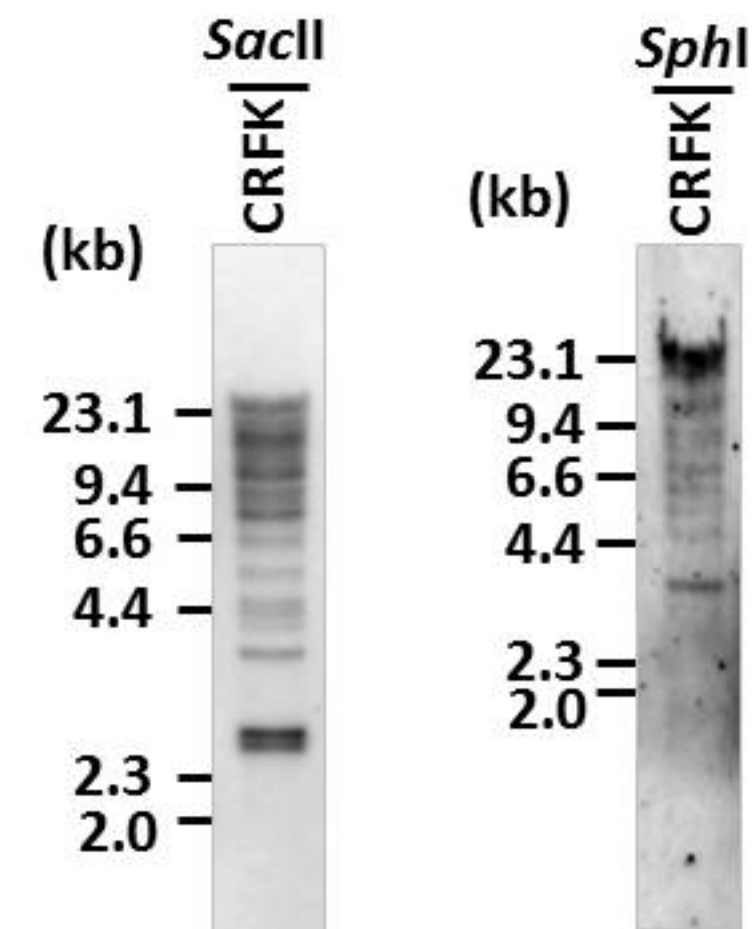
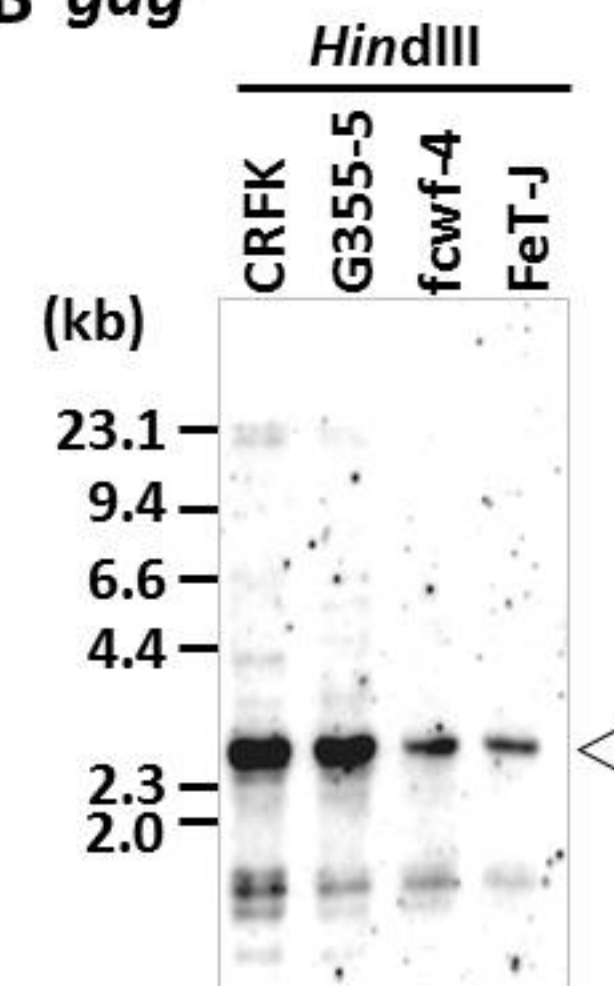
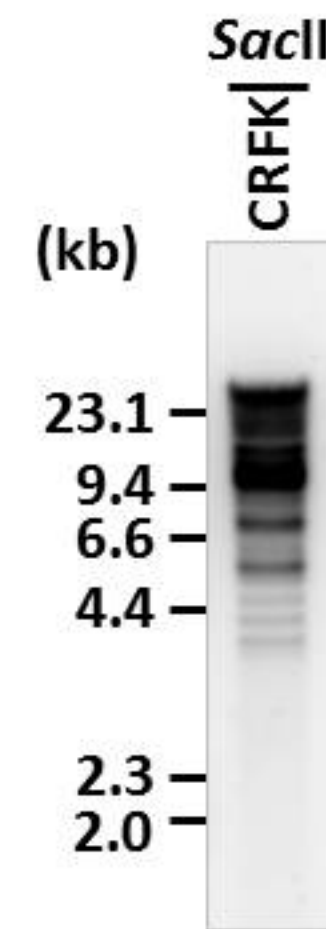
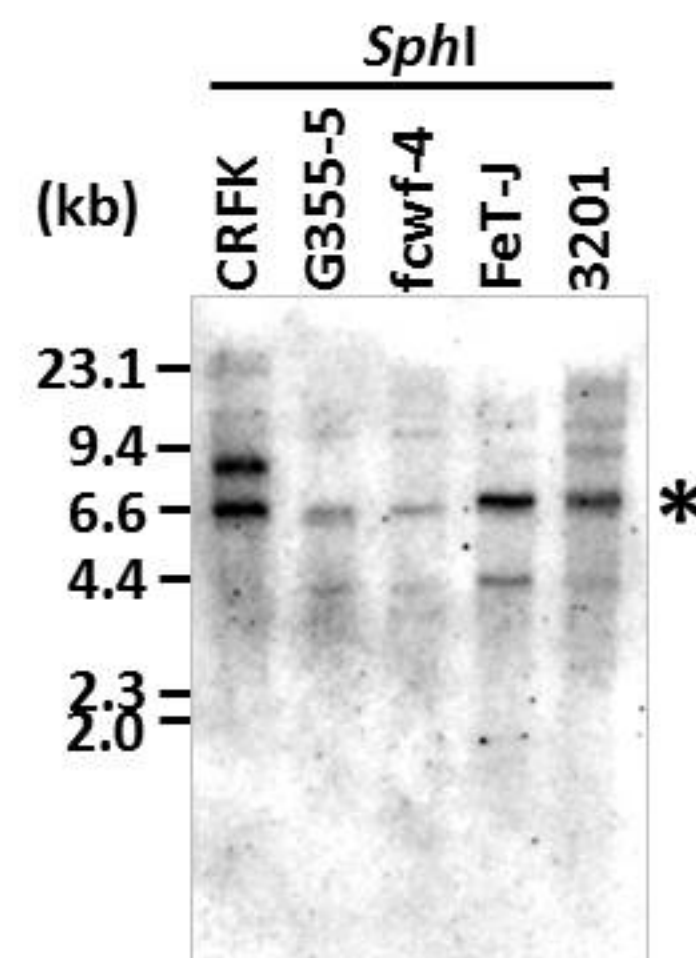
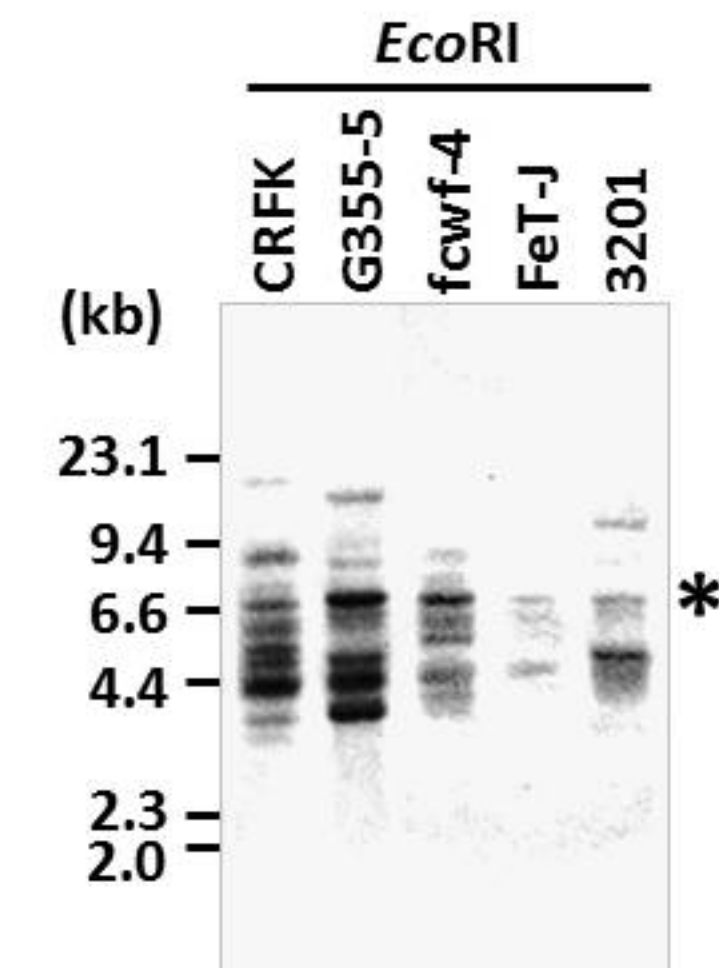
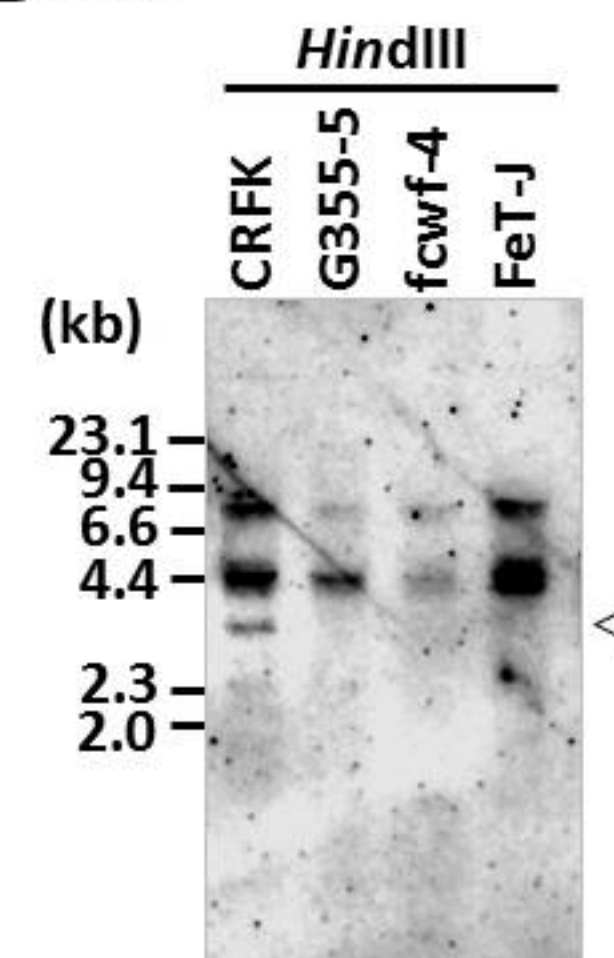
634 positions of primers used for PCR. Small letters indicate nucleotides. Capital letters in
635 parentheses indicate amino acids. Dots indicate differences between RDRV AC and the other
636 clones. Colors of dots are dependent on where are mutations from (Red, derived from A2; Blue,
637 from C1; Gray, from A2 or C1; Black, undetermined). (C) LTR structures of RD-114 virus clones
638 (pSc3c and pCRT1), RDRS A2, C1 and RDRV AC.

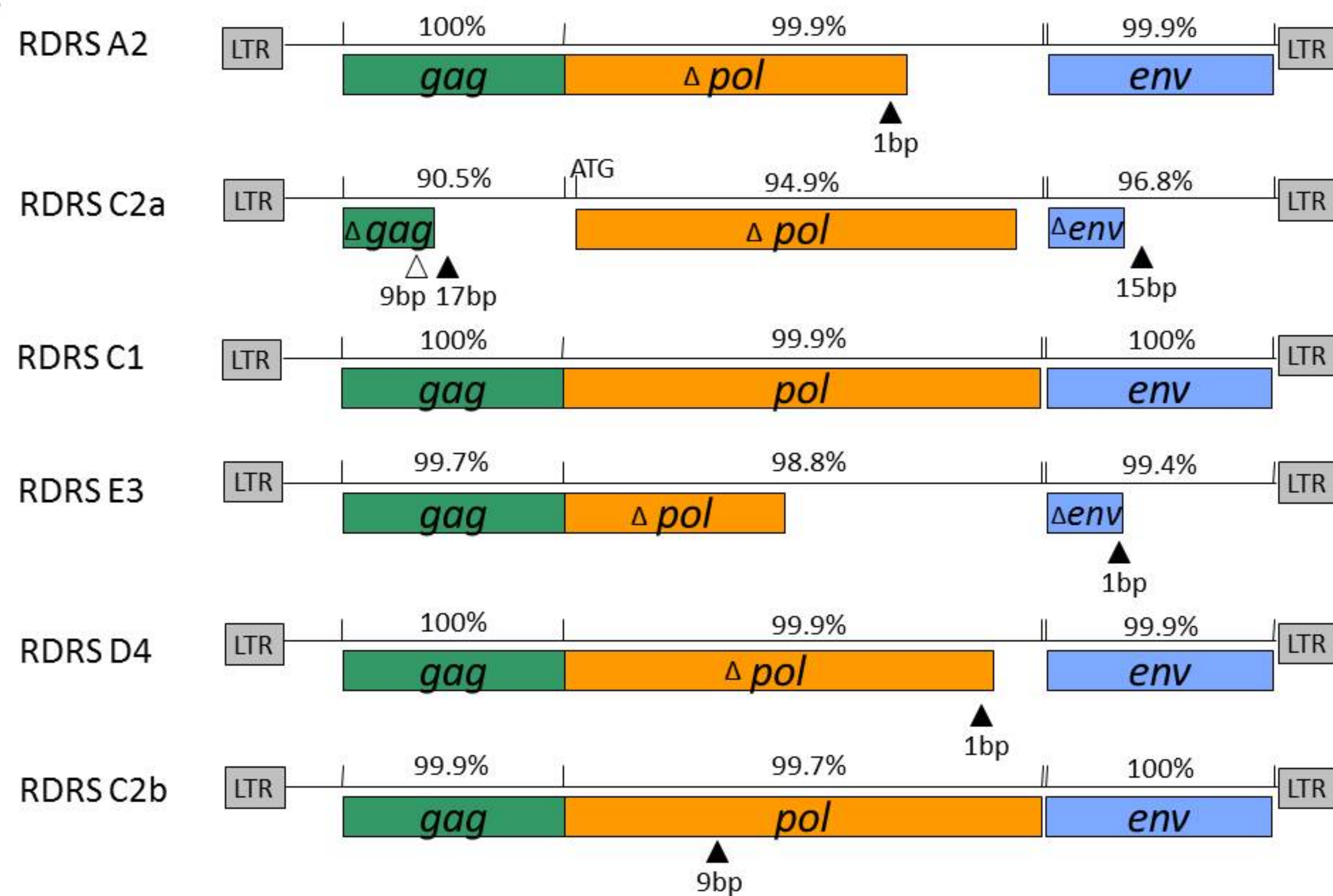
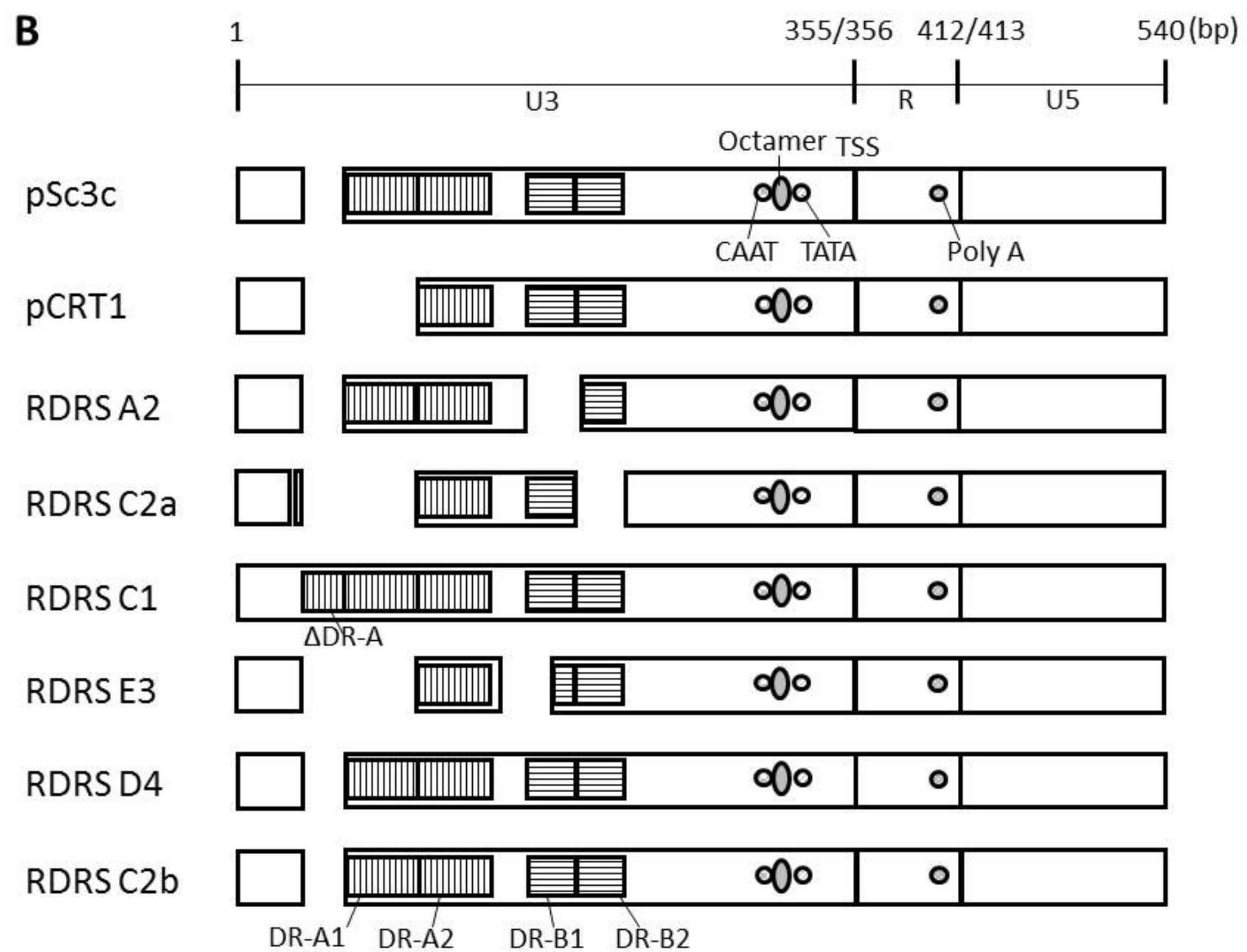
Origin	Cells	RDRS proviral integration						Infectious-type <i>env</i>	Viral production
		A2	C2a	C1	E3	D4	C2b		
kidney	CRFK	+	+	+	-	-	-	+	+
glial	G355-5	-	+	-	-	-	-	-	-
whole fetus	fcwf-4	-	+	-	-	-	-	-	-
T lymphocyte	FeT-J	-	+	-	-	-	-	-	-
lymphoma	3201	-	+	+	+	-	-	+	-
Embryo fibroblast	FER	-	+	-	-	-	+	+	+
	AH927	-	+	-	-	-	-	-	-
Total integration rates (%)		1 / 7 (14.3)	7 / 7 (100)	2 / 7 (25.0)	2 / 7 (28.6)	0 / 7 (0)	1 / 7 (14.3)	3 / 7 (42.9)	2 / 7 (28.6)

Table 1: Population of RDRS proviruses in feline cell lines.

Provirus	Chromosome	TSD sequence	Locus of TSD
RDRS A2	A2	TGATTT	153,290,986-153,290,991
RDRS C2a	C2	U.D.	U.D.
RDRS C1	C1	TGGTAT	59,327,845-59,327,850
RDRS E3	E3	TGATCT	28,850,909-28,850,914
RDRS D4	D4	TGTCTC	77,169,718-77,169,723
RDRS C2b	C2	GGCAGG	111,337,840-111,337,845

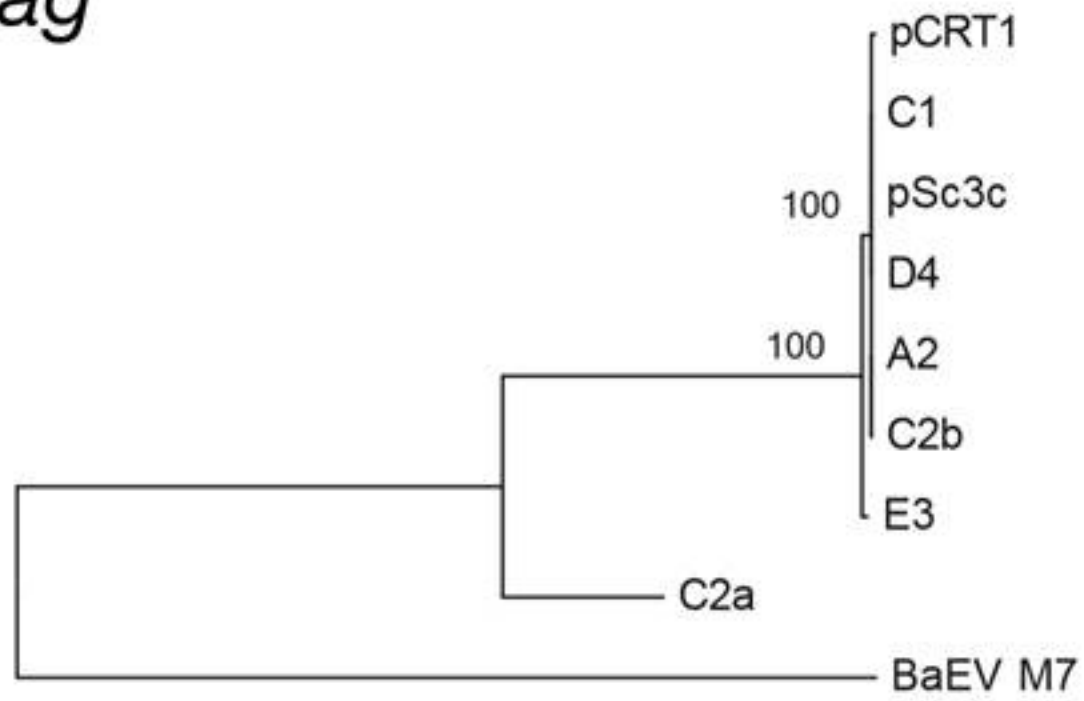
Table 2: RDRS proviral target site duplications (TSDs). Flanking 6-bp TSD sequences and loci are shown. U.D. is undetectable.

A**B gag****C pol****D env**

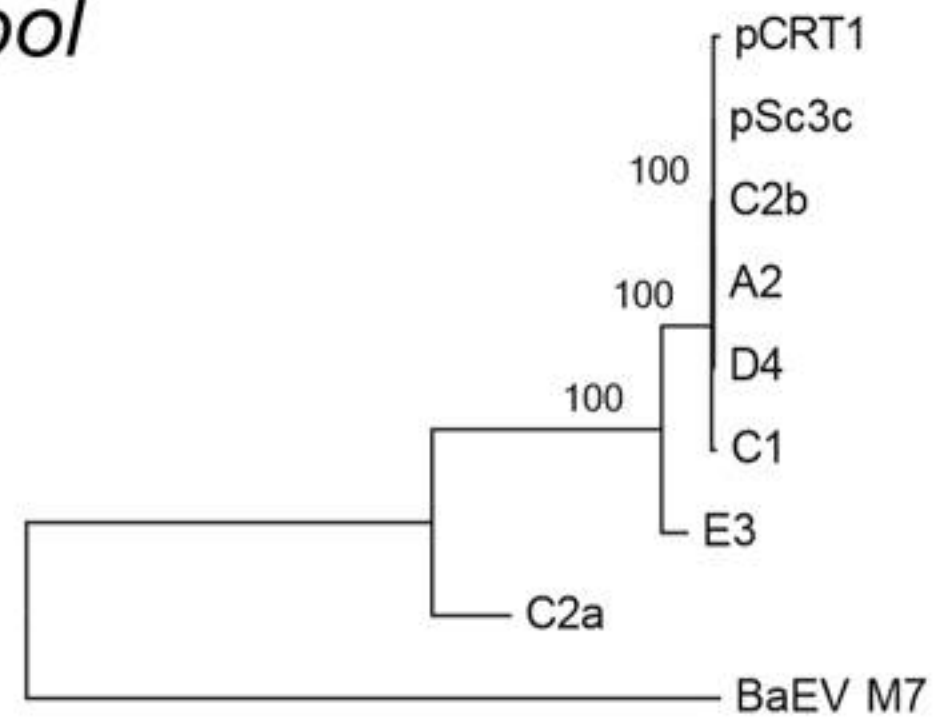
A**B****C**

Provirus	No. of differences between 5' and 3'LTRs	Homology between 5' and 3'LTRs (%)	Divergence between 5' and 3'LTR (%)	Length of LTR (bp)		Estimated integration time (MYA)
				5'	3'	
RDRS A2	0	100	0	449	449	<0.2
RDRS C2a	8	98.1	1.9	416	418	1.6
RDRS C1	1	99.8	0.2	504	504	0.2
RDRS E3	0	100	0	404	404	<0.2
RDRS D4	0	100	0	482	482	<0.2
RDRS C2b	0	100	0	482	482	<0.2

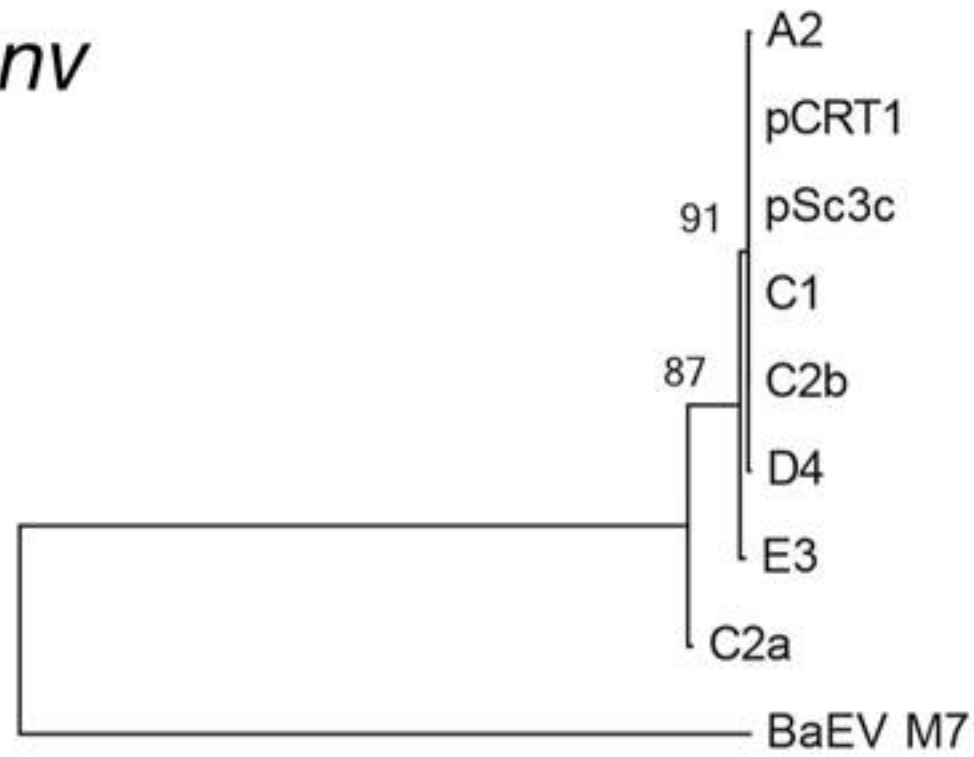
gag



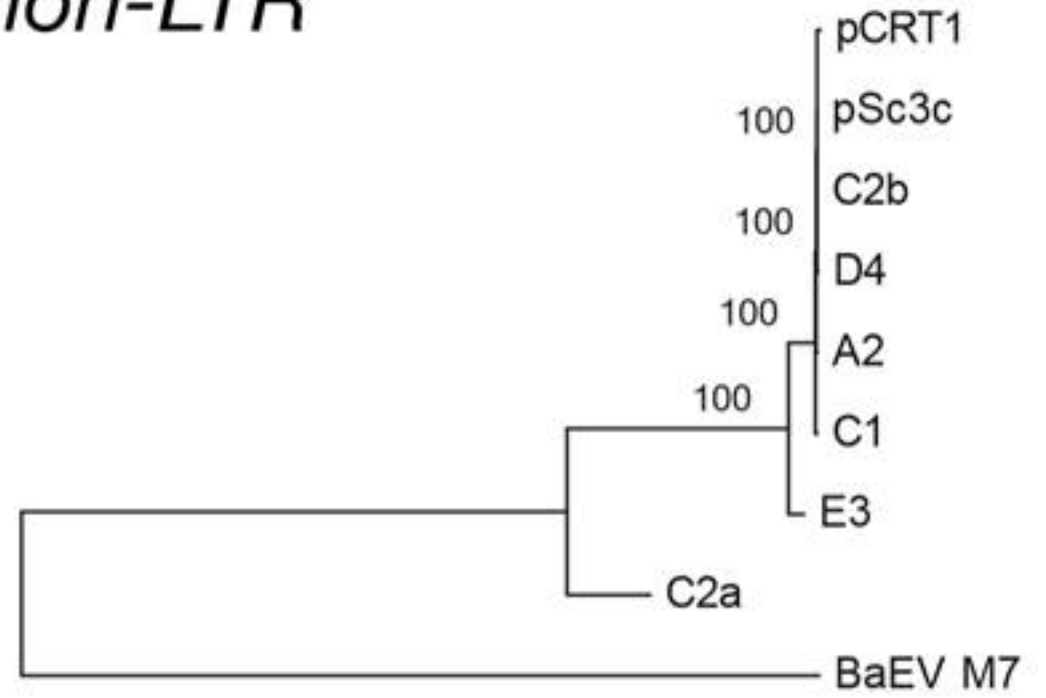
pol



env



non-LTR



A

Region	Cat Breed	RDRS proviral integration rates (%)						Infectious-type <i>env</i>
		A2	C2a	C1	E3	D4	C2b	
North America	American Curl	0 / 4 (0.0)	4 / 4 (100)	1 / 4 (25.0)	2 / 4 (50.0)	0 / 4 (0.0)	0 / 4 (0.0)	2 / 4 (50.0)
	American Shorthair	0 / 4 (0.0)	4 / 4 (100)	0 / 4 (0.0)	1 / 4 (25.0)	0 / 4 (0.0)	0 / 4 (0.0)	5 / 6 (83.3)
	Bengal	0 / 2 (0)	2 / 2 (100)	0 / 2 (0)	1 / 2 (50.0)	0 / 2 (0)	1 / 2 (50.0)	1 / 2 (50.0)
	Exotic	0 / 4 (0.0)	4 / 4 (100)	1 / 4 (25.0)	0 / 4 (0.0)	0 / 4 (0.0)	0 / 4 (0.0)	2 / 4 (50)
	Maine Coon Cat	0 / 1 (0)	1 / 1 (100)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)	0 / 2 (0)
	Total	0 / 15 (0)	15 / 15 (100)	2 / 15 (13.3)	4 / 15 (26.7)	0 / 15 (0)	1 / 15 (6.7)	10 / 18 (55.6)
Europe	British Shorthair	0 / 1 (0)	1 / 1 (100)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)
	European Shorthair	0 / 2 (0)	2 / 2 (100)	0 / 2 (0)	2 / 2 (100)	2 / 2 (100)	0 / 2 (0)	2 / 2 (100)
	Munchkin	0 / 1 (0)	1 / 1 (100)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)
	Russian Blue	0 / 5 (0)	5 / 5 (100)	0 / 5 (0)	0 / 5 (0)	0 / 5 (0)	0 / 5 (0)	2 / 6 (33.3)
	Total	0 / 9 (0)	9 / 9 (100)	0 / 9 (0)	2 / 9 (22.2)	2 / 9 (22.2)	0 / 9 (0)	4 / 10 (40.0)
North America x Europe	Himalayan	0 / 1 (0)	1 / 1 (100)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)
	Ragdoll	0 / 1 (0)	1 / 1 (100)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)	0 / 3 (0)
	Scottish Fold	0 / 17 (0)	17 / 17 (100)	0 / 17 (0)	0 / 17 (0)	0 / 17 (0)	0 / 17 (0)	12 / 17 (70.6)
	Mix breed (USA, UK)	0 / 4 (0)	4 / 4 (100)	0 / 4 (0)	0 / 4 (0)	0 / 4 (0)	0 / 4 (0)	0 / 4 (0)
	Total	0 / 23 (0)	23 / 23 (100)	0 / 23 (0)	0 / 23 (0)	0 / 23 (0)	0 / 23 (0)	12 / 25 (48.0)
Asia	Singapura	0 / 1 (0)	1 / 1 (100)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)
	Tonkinese	0 / 1 (0)	1 / 1 (100)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)	0 / 1 (0)
	Calico cat (Mi-ke)	N. A.	N. A.	N. A.	N. A.	N. A.	N. A.	0 / 1 (0)
	Mongrel	0 / 20 (0)	20 / 20 (100)	0 / 20 (0)	0 / 20 (0)	0 / 20 (0)	0 / 20 (0)	3 / 77 (3.9)
	Total	0 / 22 (0)	22 / 22 (100)	0 / 22 (0)	0 / 22 (0)	0 / 22 (0)	0 / 22 (0)	3 / 80 (3.8)
Middle East	Abyssinian	0 / 6 (0)	6 / 6 (100)	0 / 6 (0)	1 / 6 (16.7)	0 / 6 (0)	0 / 6 (0)	1 / 6 (16.7)
	Persian	0 / 2 (0)	2 / 2 (100)	0 / 2 (0)	0 / 2 (0)	0 / 2 (0)	0 / 2 (0)	2 / 2 (100)
	Total	0 / 8 (0)	8 / 8 (100)	0 / 8 (0)	1 / 8 (12.5)	0 / 8 (0)	0 / 8 (0)	3 / 8 (37.5)
Overall Total		0 / 77 (0)	77 / 77 (100)	2 / 77 (2.6)	7 / 77 (9.1)	2 / 77 (2.6)	1 / 77 (1.3)	42 / 141 (29.8)

B

Infectious-type *env*

6950 6960 6970 6980 6990 7000

TCTTGC GATACCCACTCCCTCTTTAACCTACTCCCTAGCAGACTCCCTAGCGAATGC

|||||

AGAACGC **TATGGGTGAGGGAGAAATTGGATGAGGGATCGTCTGAGGGATCGCTTACG**

←

RDRS C2a *env*

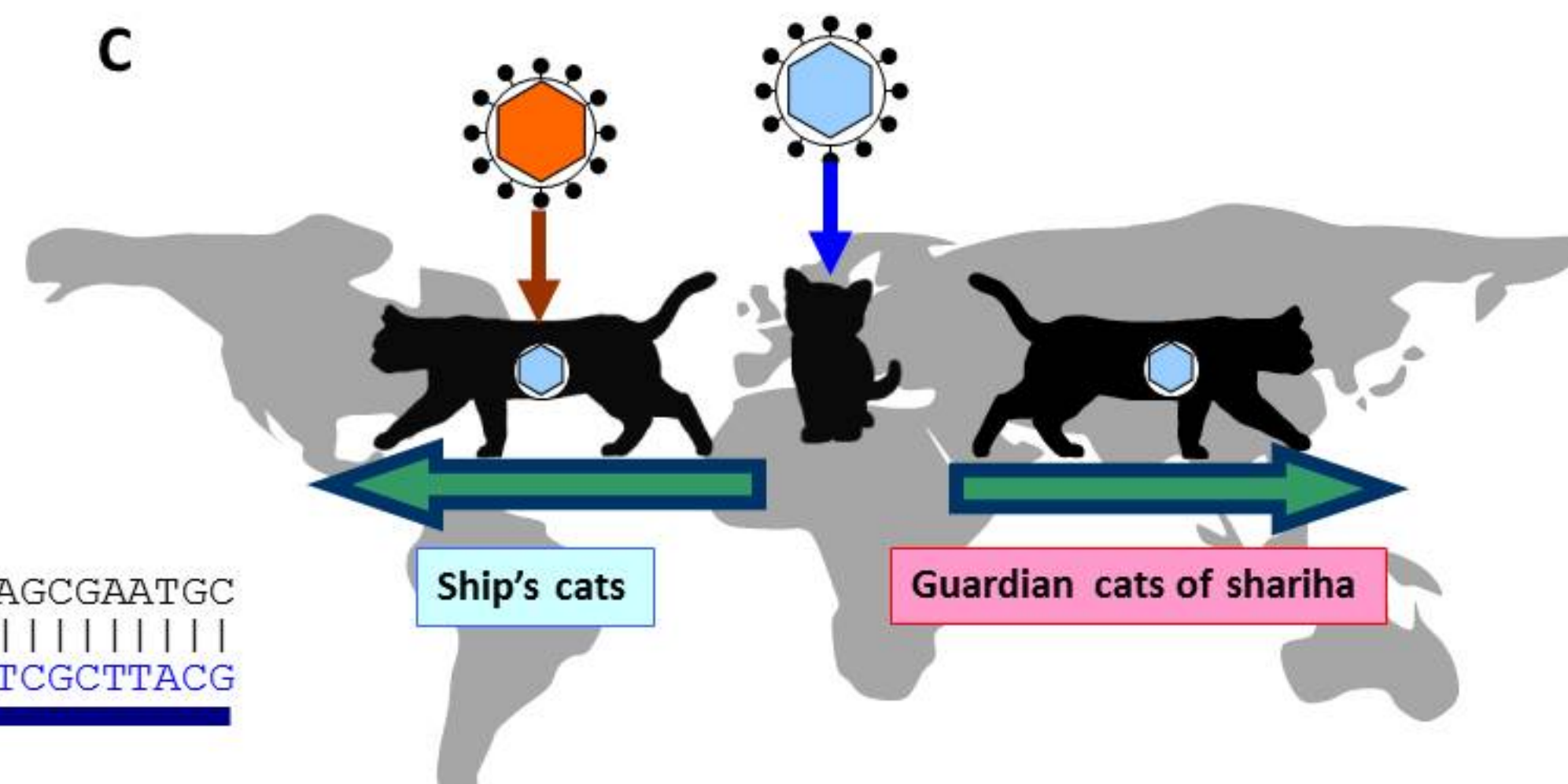
TCTTGC GATAC **CCACAATCCCTGGAC** CCACTCCCTCTTTAACCTACTCCCTAGCAGACTCCCTAGCGAATGC

|||||

AGAACGC **TATGGGTGTTAGGGACCTGGGTGAGGGAGAAATTGGATGAGGGATCGTCTGAGGGATCGCTTACG**

←

C



[illegible]

0 1 8 9 (kb)

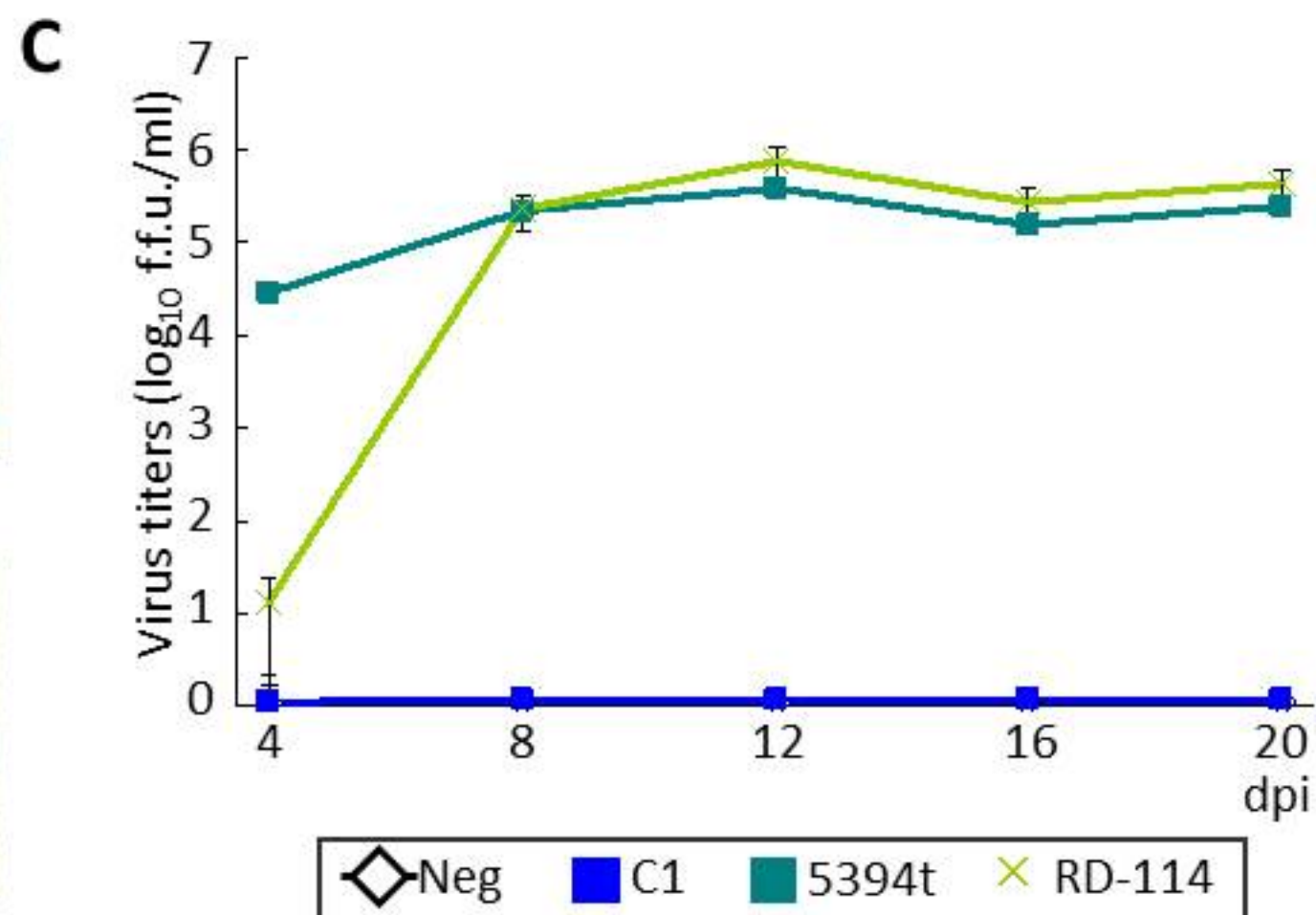
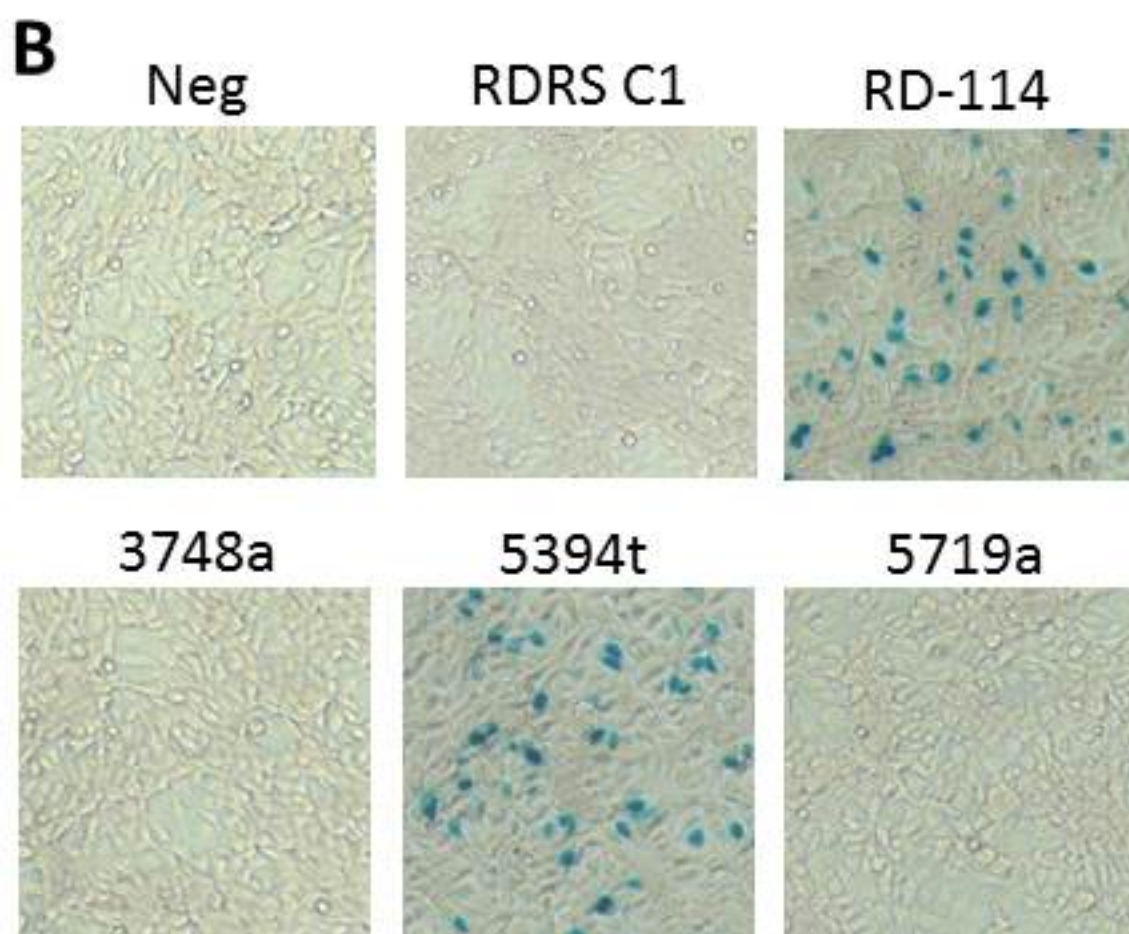
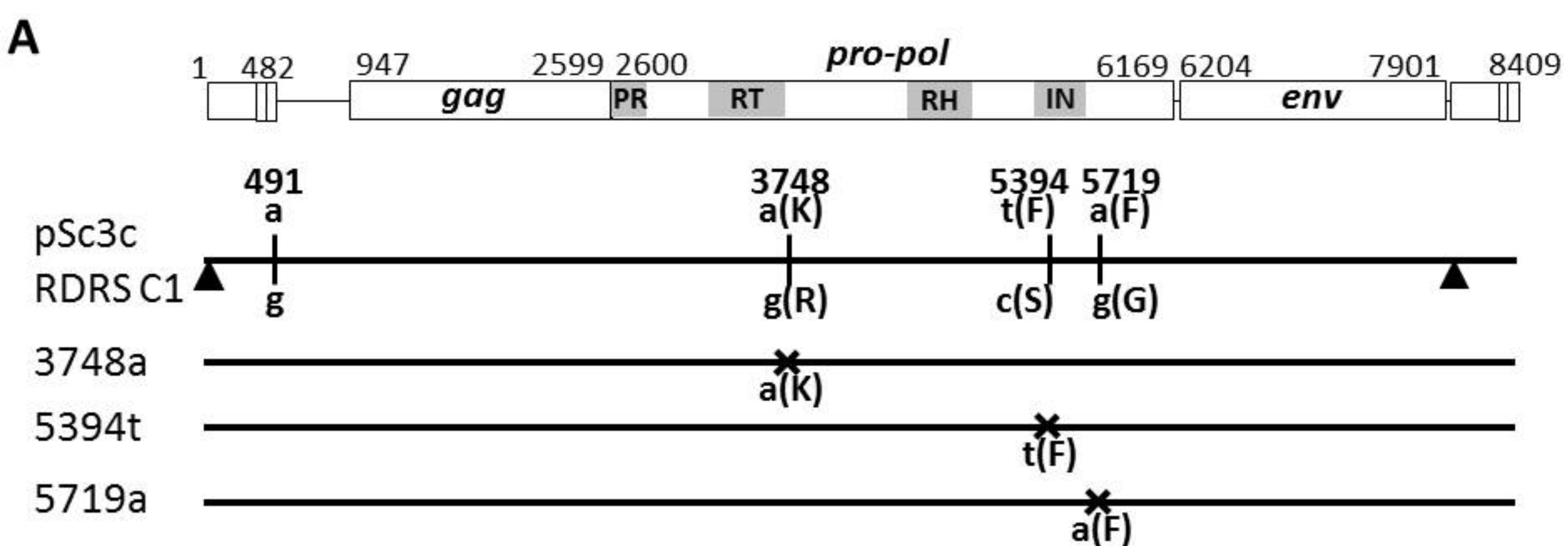
F. catus

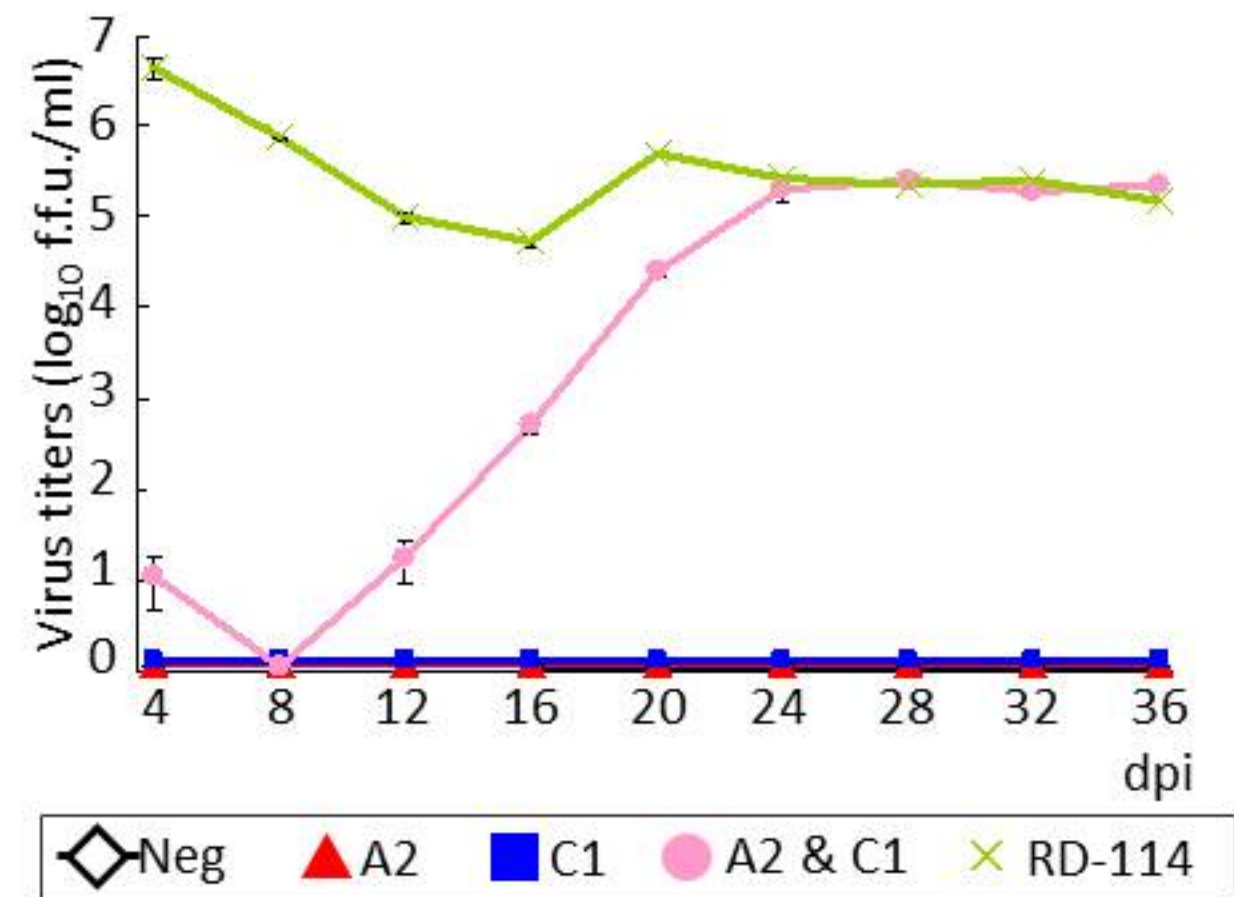
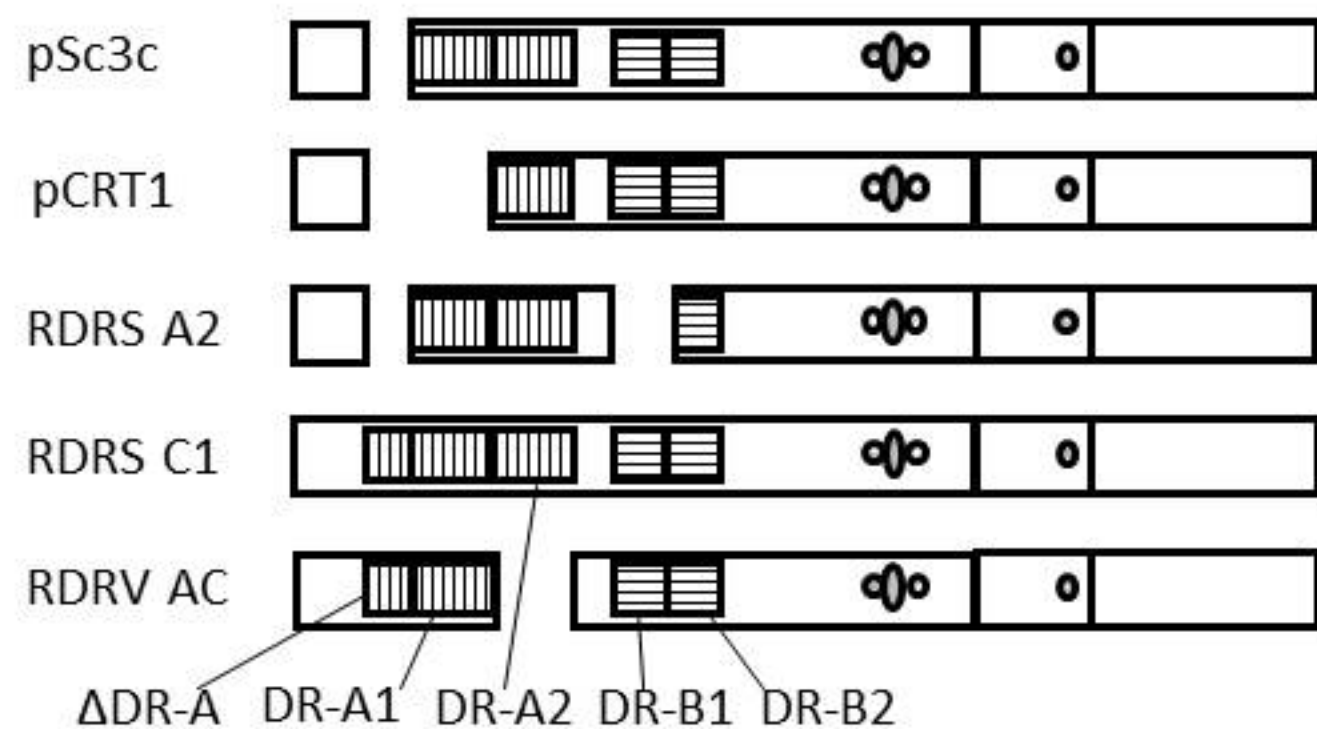
P. bengalensis

L. serval

P. uncia

P. tigris



A**C****B**