Identification and classification of feline endogenous retroviruses in the cat genome using degenerate PCR and in silico data analysis

Ning Song,¹ Hain Jo,¹ Minkyeung Choi,¹ Jin-Hoi Kim,¹ Han Geuk Seo,¹ Se-Yeoun Cha,² Kunho Seo³ and Chankyu Park¹

¹Department of Animal Biotechnology, Konkuk University, Hwayang-dong, Kwangjin-gu, Seoul, South Korea
²College of Veterinary Medicine, Chonbuk National University, Jeonju, South Korea
³Colleges of Veterinary Medicine, Konkuk University, Seoul, South Korea

The purpose of this study was to identify and classify endogenous retroviruses (ERVs) in the cat genome. Pooled DNA from five domestic cats was subjected to degenerate PCR with primers specific to the conserved retroviral propol region. The 59 amplified retroviral sequences were used for in silico analysis of the cat genome (Felis catus-6.2). We identified 219 ERV γ and β elements from cat genome contigs, which were classified into 42 ERV γ and 4 β families and further analysed. Among them, 99 γ and 5 β ERV elements contained the complete retroviral structure. Furthermore, we identified 757 spuma-like ERV elements based on the sequence homology to murine (Mu)ERV-L and human (H)ERV-L. To the best of our knowledge, this is the first detailed genome-scale analysis examining Felis catus endogenous retroviruses (FcERV) and providing advanced insights into their structural characteristics, localization in the genome, and diversity.

INTRODUCTION

Endogenous retroviruses (ERVs) are copies of exogenous retroviral genomes integrated into the host genome. They are classified into seven genera: alpha-, beta-, gamma- and epsilonretroviruses, lentiviruses and spuma-like retroviruses have complex genomes. It has been previously shown that all simple retroviruses can become endogenous in their hosts (Gifford & Tristem, 2003; Weiss, 2006). In the case of complex retroviruses, only spumaviruses have been found to have the ability to become endogenous inside diverse host species including mammals such as mice, humans, rats, rabbits, dogs, cows, elephants and aye-aye (Bénit et al., 1999; Greenwood et al., 2004; Han & Worobey, 2012a; Jo et al., 2012), coelacanth (Han & Worobey, 2012b) and birds (Bolisetty et al., 2012). However, in recent studies endogenous lentiviruses were also identified in the genomes of a few host species, including rabbits, weasel and ferrets (Cui & Holmes, 2012; Han & Worobey, 2012c; Katzourakis et al., 2007).

Once ERVs were integrated into host genomes, they co-evolved with their host genomes (McDonald, 1993). Similar retroviral sequences in diverse species indicate a cross-species transmission event (Bénit et al., 2001). Understanding the influence of ERVs on animal genome modification is an interesting issue in genome biology (Hughes & Coffin, 2001), and thus identifying and characterizing ERV-related sequences in diverse host genomes has become important.

Numerous ERV sequences from diverse vertebrate genomes have been characterized using various techniques (Gifford & Tristem, 2003). One of the earliest approaches was to stimulate the induction of virus release from normal cells after treatment with ionizing radiation, chemical mutagens, or carcinogens (Weiss et al., 1971). ERVs were also detected based on their homology to synthetic primer binding site probes (Kröger & Horak, 1987) as well as DNA hybridization and sequencing (Dunwiddie et al., 1986; Lueders & Kuff, 1980; Ono et al., 1985). Recently, sufficient data for phylogenetic analyses were successfully obtained by using PCR to directly amplify ERV sequences from the host genomic DNA of pigs (Klymiuk et al., 2002; Patience et al., 2001), sheep (Klymiuk et al., 2003), cattle (Garcia-Etxebarria & Jugo, 2010; Xiao et al., 2008b), dog (Jo et al., 2012) and other vertebrates (Herniou et al., 1998). The analysis results reported three classes of ERVs, class I (γ [C-type]), class II (β [B/D-type]), and class III (spuma-like) in those genomes.

Several endogenous retroviruses have been detected in the genome of domestic cats (Felis catus); these viruses include...
endogenous feline leukaemia virus (enFeLV) (Soe et al., 1983), RD-114, a C-type retrovirus (McAllister et al., 1972), MAC-1 primate retrovirus-related virus (Bonner & Todaro, 1979), ECE1 (Beyer et al., 1987), FeEV (van der Kuyl et al., 1999), FERVmlu1 and FERVmlu2 (Yuhki et al., 2008), and ERV-DC (Anai et al., 2012). However, detailed analysis regarding their structural variations, genome-level diversity and their relationships has not been addressed.

House cats have become a familiar companion to people since their domestication from the Asian wildcat (Felis silvestris lybica) approximately 10 000 years ago in the Middle East’s Fertile Crescent (Driscoll et al., 2007). Together with dogs, extensive veterinary attention has been given to domestic cats, more than any other animal (Mullikin et al., 2010). Various resources of feline veterinary models provide a unique opportunity to explore the genetic determinants responsible for genetic diseases, infectious disease susceptibility, behavioural and neurological phenotypes, reproduction and physiology (O’Brien et al., 2008; Online Mendelian Inheritance in Animals (OMIA), http://omia.angis.org.au/home/).

In this study, we first amplified the conserved pro/pol nucleotide sequences of ERVs in the cat genome using degenerate primers specific to the active site motifs, DTGA of protease (PR) and YMDD or YVDD of reverse transcriptase (RT) proteins. Based on the results of this analysis, we subsequently analysed ERV sequences from the current cat genome assembly. We also classified ERV elements in the cat genome to understand the diversity and evolution of FcERV-related sequences, and compared them with several previously reported feline ERV sequences.

RESULTS

Identification and classification of ERV sequences from pro/pol-specific degenerate PCR using cat genomic DNA

Degenerate PCR using cat genomic DNA and retroviral pro/pol-specific primers and subsequent cloning yielded 194 clones. Sequence analysis of the clones showed that 59 (30%) clones with a length of 0.8 to 1.1 kb had significant sequence similarity (66–99%) to other previously reported retroviral sequences, in which 55 and four sequences showed high sequence homology (98% similarity) to enFeLV-AGTT, FeLV-A, RD-114, ERV-DC10, FERVmlu1, FERVmlu2, CFERV γ1, BaEV, VuEV, TaEV, HaEV, MiEVII and MeEVII) and betaretroviruses (Mason-Pfizer monkey virus, MPMV) and also performed phylogenetic analysis using the nucleotide sequences of previously reported ERVs and representative ERVs obtained by performing degenerate PCR (Fig. 1). The clone, MJ52, was almost identical (99% similarity) to ERV-DC10, which has been previously reported to be an infectious ERV in domestic cats (Anai et al., 2012). Another clone MJ62 showed high sequence homology (98% similarity) to enFeLV-AGGAG and enFeLV-AGTT. The clone PC37 showed 93% and 92% sequence similarity to FERVmlu1 and FERVmlu2 (Yuhki et al., 2008), respectively. Only partial sequences of the pol region of ECE1 (GenBank accession number X51929) and FeEV (AF164923) were available, and the analysis showed that they were closely related to RD-114 pol (data not shown).

However, the phylogenetic relationship of a previously reported feline ERV, MAC-1 (Bonner & Todaro, 1979), could not be determined because the sequence information for this virus was not available. Therefore, the pro/pol sequences of all previously reported feline endogenous retroviruses with available nucleotide sequence information were matched with those of FcERVs from this study, and a high sequence homology (>92%) was observed.

Degenerate PCR amplified clones formed nine ERV γ and one β families (Fig. 2). The most abundant group in the gammaretroviruses was assigned as FcERV γ1, which constituted 32% of ERV sequences from degenerate PCR. The clones belonging to γ2 to γ9 families were determined according to their genetic distance from FcERV γ1 (Table 1). Although FcERV γ1 and FcERV γ2 were from the same cluster, FcERV γ2 was classified into a family different from that of FcERV γ1; the classification was such because we did not consider the extra 206 bp in the pol region of FcERV γ2 while comparing this ERV to other FcERV families during the phylogenetic analysis. If the extra sequence is taken into consideration, FcERVs γ1 and γ2 will have to be classified into different clusters (data not shown). The results of our analysis showed that FERVmlu1 and 2 are identical to FeERV γ6, RD-114 and ERV-DC10 are identical to FcERV γ7, and enFeLV is identical to FcERV γ8. Therefore, we can conclude that several previously reported cat ERVs are all included in our family classification of cat ERVs (Figs 1 and 2).

In silico identification of feline endogenous retrovirus using degenerate PCR results

For the in silico analysis of cat ERV, we performed BLASTN analysis using 59 representative sequences from degenerate PCR with the fewest nonsense mutations and thus the most closely related to potentially infectious ERVs were selected as representatives for each family and deposited into GenBank under accession numbers JF93910–JF93199.

Each representative ERV nucleotide sequence was BLASTed against the previously reported reference sequences of ERVs such as gammaretroviruses (enFeLV-GGAG, enFeLV-AGTT, FeLV-A, RD-114, ERV-DC10, FERVmlu1, FERVmlu2, CFERV γ1, BaEV, VuEV, TaEV, HaEV, MiEVII and MeEVII) and betaretroviruses (Mason-Pfizer monkey virus, MPMV).
PCR clones and three \textit{F. catus} origin ERV sequences from the Repbase against the cat genome assembly, \textit{Felis catus-6.2}. According to matching criteria described in Methods, \textit{BLAST} analysis identified 219 \textit{pro/pol}-containing elements (data not shown). Among them, 210 sequences showed close relationships to ERV \(\gamma\) and 9 to \(\beta\) elements. Only \(\gamma\) and \(\beta\) retroviruses were identified from the \textit{BLAST} analysis, indicating that the degenerate primers used in this study were effective for amplifying the \textit{pro/pol} regions of ERV \(\gamma\) and \(\beta\) elements and were not efficient for identifying the sequences of other ERV families such as spuma-like ERVs.

**Classification of ERVs in the cat genome into 42 \(\gamma\) and 4 \(\beta\) families**

To establish the family classifications for ERV elements in the cat genome, we combined the identified \textit{pro/pol} sequences from degenerate PCR and \textit{in silico} analysis. From the \textit{BLASTX} analysis, we found that there are both \textit{PR-} and \textit{RT}-containing elements, and \textit{PR-} or \textit{RT}-containing elements among the \textit{in silico} identified ERV elements (Table 2). Therefore, we separated the sequences into three different groups, \textit{PR-RT}, \textit{PR-only} and \textit{RT-only} groups, depending on the presence of motifs, and performed ERV classification analysis. Both phylogenetic analysis and sequence similarity comparison were performed. The criteria for ERV classification were the same as those used for degenerate PCR sequences.

For ERV \(\gamma\) elements, 245 \textit{PR-RT} group elements (92.5 \%), including 190 clones with \textit{in silico} and 55 with degenerate PCR origins, were subjected to family classification. The analysis showed that the sequences were clustered into 33 groups, \textit{FcERV \(\gamma_1\) to \(\gamma_{33}\)}, with 24 new families in addition to \textit{FcERV \(\gamma_1\)} to \(\gamma_9\) from degenerate PCR products. \textit{FcERV \(\gamma_8\)} family sequences were only present in sequences from

---

**Fig. 1.** Phylogenetic analysis of ten representative \textit{F. catus} endogenous retroviruses (\textit{FcERVs}) generated from degenerate PCR compared with previously identified retroviruses. \textit{FcERV \(\gamma\) and \(\beta\)} families clustered with gammaretroviruses (enFeLV-GGAG, enFeLV-AGTT, FeLV-A, RD-114, ERV-DC10, FERVmlu1, FERVmlu2, CIERV \(\gamma_1\), BaEV, VuEV, TaEV, HaEV, MiEVII and MeEVII) and betaretroviruses (MPMV). A total of 930 bp of common sequences corresponding to the retroviral \textit{pro/pol} region were used for analysis. Bootstrap values (>50) from 1000 replicates are indicated at branch nodes. Asterisks indicate representative sequences.
using pooled genomic DNA from five cats with unknown breed information, while the cat genome assembly used for the in silico analysis was constructed from a female Abyssinian cat. The PR-motif-only group contained just two elements (0.75%) and was separated into two families, FcERV γ34 and FcERV γ35. The RT-motif-only group contained 18 ERV γ elements (3.02%) and formed seven families, FcERV γ36 to FcERV γ42. In the cat genome, relative to other genomes, the number of ERVs in each family was very limited. Among the 24 new ERV families of the PR-RT group identified through in silico analysis, most had only one or two ERV elements; FcERV γ10, γ11, γ13, γ14, γ16, γ17, γ18, γ19, γ20, γ21, γ22, γ23, γ24, γ29, γ31, γ32 and γ33 had single ERV elements, and FcERV γ12, γ25, γ27, γ28 and γ30 had two elements each. For PR-only ERVs, each family contained only one ERV element, and for RT-only ERVs, FcERV γ37 and FcERV γ42 contained only one element each.

For ERV β, we identified a total of 13 elements from both degenerate PCR (n=4) and in silico analysis (n=9). ERV β elements in cats were classified into four families, from FcERV β1 to β2 (PR-RT groups, Fig. 3) and from FcERV β3 to β4 (RT-only group). The sequence similarities between representative sequences of FcERV β1 and β2 elements and of FcERV β3 and β4 were 54.2% and 59.1%, respectively.

**Structural analysis of full-length FcERV γ and β elements**

Identification of long-terminal repeat (LTR) sequences for each element is required for detailed structural characterization of candidate ERV elements identified by the presence of pro/pol sequences in the contigs. We aligned the sequences corresponding to up to 10 kb regions from both 5' and 3' ends of the pro/pol sequences in the contigs using NCBI BLAST2SEQ. Based on the analysis of 219 ERV elements associated with the cat genome contigs, we identified 99 ERV γ and five β elements with LTRs at both the 5' and 3' ends (Tables S1 and S2, available in JGV Online). These candidate ERV elements were further examined using NCBI BLASTX to identify sequences corresponding to essential retroviral genes, gag, pro, pol and env. For γ ERVs, 49.5% (n=49) of the sequences contained the complete gag-pro-pol-env structure among the 99 FcERVs with both 3' and 5' LTRs, 48.53% (n=48) contained the gag-pro-pol structure, and the remaining sequences contained the pro-pol-env or pol-env structure (Table S3). For β ERVs (n=5), two showed the complete structure, two had gag-pro-pol, and one had the gag-pol structure (Table S4).

To estimate the total number of FcERV γ and β elements in the cat genome assembly (Felis catus-6.2), we calculated the mean size (6751 bp) of FcERVs from the 104 full-length sequences and multiplied the value by the total number of identified FcERV γ and β elements in the cat genome. We estimated that the total amount of FcERV γ

---

**Fig. 2.** Phylogenetic analysis and family classification of 59 PR- and RT-containing ERV sequences from degenerate PCR. A neighbour-joining tree constructed using nucleotide sequences of 59 pol sequences gave rise to ten cat ERV families named FcERV γ1 to γ9 and FcERV β1. A total of 1020 bp with common sequences corresponding to the retroviral pol region were used for analysis. Bootstrap values (≥50) from 1000 replicates are indicated at branch nodes. Asterisks indicate representative sequences.

---

degenerate PCR, suggesting that ERVs present among different individuals or breeds can differ according to their DNA sources. In this study, degenerate PCR was performed...
and β elements in *Felis catus*-6.2 is 0.06 % of the cat genome or 1.48 Mb.

### Identification of spuma-like ERV elements in the cat genome

To address the presence of class III ERVs in the cat genome, we performed BLAST analysis against the cat genome using spuma-like elements in humans and mice, HERV-L (X89211) and MuERV-L (Y12713). BLASTX analyses using the two representative spuma-like ERVs against the cat genome assembly *Felis catus*-6.2 identified 757 matches (data not shown). Among these, 302 matches were *pro-pol*-containing elements. It has been reported that old ERV elements such as spuma-like ERVs show extreme sequence divergence and thus their classification is difficult (Blomberg et al., 2009; Jo et al., 2012). Consistent with this, we were unable to establish meaningful clustering patterns for the family classification of spuma-like FcERV (FcERV-L) elements, although we used slightly less stringent criteria when comparing the classifications of FcERV γ and β elements (data not shown). In most cases, sequence identities were lower than 70 % and therefore they were separated into single-element clusters. To approximately estimate the total amount of the FcERV-L elements in the cat genome (*Felis catus*-6.2), we multiplied the mean size of HERV-L and MuERV-L (6530 bp) by the number of identified FcERV-L elements, resulting in 4.94 Mb or approximately 0.2 % of the cat genome.

### Conservation of amino acid motifs in the PR and RT regions of FcERV elements

We analysed conserved PR and RT motifs of 104 structurally complete FcERV γ and β elements (LTR-gag-pro-pol-env-LTR) from the *in silico* analysis (Tables S3 and S4). FcERVs γ5, γ25 and γ31 showed complete conservation of both PR (DTGA) and RT (LPQGFKN and YVDD) motifs. FcERV γ1, γ4, γ6, γ26, γ27 and γ28 also showed relatively intact motif conservation with slight variations. However, the remaining FcERV γ elements showed relatively significant sequence variations. Among FcERV β, FcERV β1 showed the highest degree of motif sequence conservation with intact PR (DTGA) and RT (YMDD) motifs, although no β ERVs showed intact conservation of LPQGFKN sequences of the RT motif (Table S4). Therefore, the sequence variation of conserved PR and RT motifs of ERVs in the cat genome appears to be higher than that in other species such as dogs (Jo et al., 2012). It is less likely that these variations are the result of sequencing artefacts because the same amino acid variations are identified from multiple FcERV elements.

### Estimated ERV integration time in the cat genome

The nucleotide substitution rate per year for pseudogenes and non-coding regions in primates was calculated to be 2.3–5.0 × 10−9, and has been used to estimate retroviral integration time in mammals (Garcia-Etxebarria & Jugo, 2010; Jo et al., 2012; Tönjes & Niebert, 2003; Xiao et al., 2008a). Sequence identities of <75 % and length difference of >5 bp between 5′ and 3′ LTRs were excluded to minimize the outlier effects.

First, we calculated the evolutionary time of the two most abundant ERV families, FcERV γ1 and FcERV γ6. We
estimated that the age of FcERV γ1 (n=18) ranges from 8.2 to 30 million years (MY) and that of γ6 (n=35) ranges from 3.8 to 16 MY (Table 3). To evaluate the age distribution of all available FcERV elements in the cat genome, we divided 99 γ FcERV elements into three groups depending on their 5' and 3' LTR sequence identities. ERVs with LTR sequence identities between 85 and 95% composed the largest group (61%, n=47), followed by the high similarity group (95 to 100%) and low similarity group (75 to 85%) with frequencies of 31.2% (n=24) and 0.78% (n=6), respectively. Therefore, proviral integration time for FcERV γ and β elements was estimated to be between approximately 2.3 and 41.7 MY (Table S5).

**DISCUSSION**

In this study, we identified and classified ERVs in the cat genome. Although there has been a brief report on a genome-level analysis of feline ERVs among reports on the cat genome project (Pontius et al., 2007), only a few types of feline ERV sequences have been reported (Anai et al., 2012; Beyer et al., 1987; Bonner & Todaro, 1979; Haapala et al., 1985; McAllister et al., 1972; Soe et al., 1983; van der Kuyl et al., 1999), and a detailed analysis on feline ERVs at the genome level has not been conducted. Therefore, we aimed to identify retroviral pro/pol-containing sequences in the cat genome and group them into FcERV γ, β and -L (spuma-like) families according to the genera classification strategy for retroviruses.

An interesting result of the ERV analysis in the cat genome was that the number of γ retroviral families was much higher in cats (n=42 families) than in dogs (n=17; Jo et al., 2012). Taking into account that the analysis methods for cats and dogs are identical, the diversity of FcERV γ elements is much higher than that in other species. This is also consistent with the higher sequence variation of the conserved PR and RT motifs in cat ERVs (Table S3). However, their integration time measured based on LTR diversity was in a similar range for ERVs in dog (Jo et al., 2012) and cat genomes. Additionally, we tried to compare the family number of ERV γ elements in cats with those of other species which have reported genome-level analysis results on ERV elements. However, the lack of whole genome sequence information during the previous studies (Klymiuk et al., 2002, 2003; Löwer et al., 1996; Patience et al., 2001; Xiao et al., 2008a) and differences in analysis methods (Garcia-Etxebarria & Jugo, 2010) prevented us from further evaluating species differences in ERV γ diversity.

![Phylogenetic analysis and family classification of 11 PR- and RT-containing ERV β sequences from degenerate PCR (names starting with ME) and in silico analysis (starting with FcERV β). A neighbour-joining tree constructed using nucleotide sequences of 11 pro/pol sequences gave rise to two cat ERV families named FcERV β1 and β2. A total of 890 bp of common sequences corresponding to the retroviral pro/pol region were selected after multiple alignments and used for analysis. Bootstrap values (>50) from 1000 replicates are indicated at branch nodes.](image-url)
ERV-resembling sequences or repeat elements. Dogs have a close phylogenetic relationship with cats. The estimated value for the total ERV elements in the canine genome from our previous study under the same criteria was equal to approximately 0.1% of the dog genome (Jo et al., 2012).

Considering the close evolutionary relationship between cats and dogs as members of the mammalian order Carnivora, we predicted that a relationship exists between ERV repertoires in both animals. Therefore, we compared the sequence similarity (data not shown) and the phylogenetic relationships (Fig. 4) of the representative sequence of each ERV family (PR and RT groups) between cats and dogs, including 33 cat and eight dog ERV families (Jo et al., 2012). Only one dog ERV family, CfERV γ1, showed significant inter-species sequence similarity (85.1%) with FcERV γ26, suggesting that this element may have originated from the same ancestor of cats and dogs. However, because the integration time of FcERV γ26 was relatively recent (3.5–14 MY ago) according to our analysis based on the nucleotide substitution rate of LTRs and the divergence time between dogs and cats is estimated to be 55 MY (Pontius et al., 2007), the integration of FcERV γ26-like ERV elements into both the dog and cat genome appears to have occurred independently and not through an vertical transfer.

It has been reported that HERV-W or syncitin 1 plays an important role in placentogenesis (Mi et al., 2000; Ruebner et al., 2012). Pontius et al. (2007) reported five FERV families (FERV-1 to -5) and mentioned the sequence similarity of FERV-4 to HERV-W (syncitin 1). However, we were not able to identify HERV-W or syncitin 1 candidates from the current cat genome assembly. The closest match to HERV-W from the BLAST analysis against the cat genome showed only 15% and 74% of sequence coverage and identity to HERV-W, respectively (data not shown).

There is evidence of the presence of recombination between endogenous and exogenous feline ERVs (Anai et al., 2012; Beyer et al., 1987; Chakrabarti et al., 1994; Sheets et al., 1993; Stewart et al., 1986, 2011). However, we only focused on the classification of FcERV based on the pro/pol region and did not analyse the recombinant forms of ERVs in the cat genome.

Here, we analysed retroviral pro/pol sequence identity and classified ERV elements in the feline genome into 42 ERV γ and four β families, and found that all the previously reported ERVs were included in specific family groups. The results of this study can be used as a framework for future studies regarding the influence of ERVs in shaping and modifying the cat genome both structurally and functionally.

**METHODS**

**Degenerate PCR amplification and cloning of F. catus endogenous retroviral sequences.** Genomic DNA was prepared from ~0.5 g of tissue samples using a simple lysis method (Miller et al., 1988). An equal amount of DNA from the muscle tissues of five domestic cats obtained from the Conservation Genome Resource Bank for Korean wildlife (http://www.cgrb.org) was pooled together to be used as the template DNA for PCR. PCR parameters consisted of 2 min at 80 °C followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 43 °C for 30 s, and extension at 74 °C for 1 min, and finally one cycle at 94 °C for 30 s, 43 °C for 3 min, and 74 °C for 10 min. Amplification was performed in 25 µl reactions containing 40 pmol of each primer, 200 µM dNTPs, PCR buffer [10 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM MgCl2], 100 ng of genomic DNA, and 2 U of Taq polymerase (SuperTherm; Roche). Six pairs of degenerate primers (Herniou et al., 1998; Klymiuk et al., 2002, 2003; Tristem et al., 1996) included two forward primers, PRO (5'-GTK TTI KTI GAY ACI GGI KC-3') and MLV-PRO (5'-YTI KTI GAY ACI GGI GCT SA-3'), and three reverse primers, CT (5'-AGI AGG TCR TCI ACR TAS TG-3'), IO (5'-ATI AGI AKR TCR TCI ACR TA-3'), and EM (5'-ATI AGI AKR TCR TCC ATR TA-3'). PCR products were gel purified using the QIAquick Gel Extraction kit (Qiagen) and ligated into the pGEM-T Easy vector (Promega). Ligation products were electroporated into DH10B cells (Invitrogen) using MicroPulser (Bio-Rad). Transformed bacteria were plated onto agar containing 50 µg ml⁻¹ ampicillin, 40 mg ml⁻¹ X-Gal, and 100 mM isopropyl thiogalactopyranoside (IPTG).

**Sequencing of cloned DNA inserts.** Plasmids were isolated using a GeneAll Exprep Plasmid SV kit (GeneAll Biotechnology). Cloned pro/pol regions were sequenced using the ABI PRISM BigDye Terminator Cycle Sequencing kit (Applied Biosystems) with SP6 (5'-GAT TTA GGT GAC ACT ATA G-3') and T7 (5'-TAA TAC GAC TCA CTA TAG GG-3') universal sequencing primers in the forward and reverse directions, respectively. The full sequences of clone inserts were obtained by overlapping forward and reverse sequencing results. Sequencing results with ambiguous peaks in the chromatograms or unique polymorphisms were confirmed by resequencing the same clone. Sequences were adjusted by eliminating vector and primer sequences.
In silico identification of the *F. catus* endogenous retrovirus. The latest *F. catus* genome assembly, *F. catus*-6.2 (Lindblad-Toh et al., 2005), was obtained from the NCBI Genome database (http://0-www.ncbi.nlm.nih.gov/elia.tmu.edu.tw/nuccore/AANG00000000) and used for in silico identification of FcERVs. The genome assembly was constructed with approximately 2.45 Gb of sequences, covering approximately 90% of the genome. A previous phylogenetic analysis of diverse RT-containing elements showed that sequences of retroviral origin can be differentiated from those of non-retroviral elements such as retrotransposons (Xiong & Eickbush, 1990). We performed BLAST analysis against the cat genome using ERV sequences from three different sources; 59 *prov* sequences (814–943 bp) obtained from the degenerative PCR, three *F. catus* origin ERV sequences from the Repbase (http://www.girinst.org/repbase/index.html), and two spuma-like ERV elements (MuERV-L and HERV-L). Redundant or severely mutated sequences were removed from the BLAST matches and individual ERV candidate sequences were reiteratively BLASTed against *F. catus*-6.2 until no new ERV-related sequences were detected. Matching criteria for the in silico analysis using BLAST analysis for identifying η, β and spuma-like ERVs were 60% nucleotide sequence identity with at least 30% coverage or e-value of <10^-25.

Multiple sequence alignments and phylogenetic analysis. Multiple sequence alignments of the nucleotide or amino acid sequences were constructed using the CLUSTALW2 program (Chenna et al., 2003). Phylogenetic analyses were performed using MEGA4.1 (Tamura et al., 2007). A neighbour-joining tree was built, and bootstrap values were obtained from 1000 replicates. To define families, nucleotide sequence similarity, which was estimated by pairwise comparisons using the BioEdit program (Hall, 1999), was used. Sequences with >80% sequence identity were grouped into the same family.

Substructural characterization of feline endogenous retroviruses. For structural characterization of candidate ERV elements identified by the presence of *prov* in the sequences, sequences corresponding to 7000–10,000 bp regions from both the 5' and 3' ends of the *prov* sequences were aligned against one another using NCBI BLAST2SEQ to evaluate the presence of LTRs (Jo et al., 2012). The typical features of LTRs, including the presence of dinucleotide inverted repeat (TG-GA) (Baillie & Wilkins, 2001; Craven et al., 1995), polyadenylation signal, a TATA box and a CCAAT box (Boonyaratanakornkit et al., 2004; Harada et al., 1987), were examined. We considered sequences a minimum of 100 bp in length as possible LTR sequences. Further confirmation of ERV substructures was conducted by structural recognition of gag, prov, pol and env genes using BLAST analysis. Conserved amino acid motifs of the PR and RT regions for each full-length ERV element were determined using NCBI BLASTX.

Estimation of LTR integration time. Both 3' and 5' LTRs of ERVs were identical in their sequences at the site of provirus formation
and without selective pressure for both LTRs independently accumulating mutations over time (Dangel et al., 1995; Johnson & Coffin, 1999; Lebedev et al., 2000). Therefore, the age of a provirus can be estimated by comparing sequences between flanking 5’ and 3’ LTRs. Relative evolutionary time was calculated as $T = D/2r$ (T, evolutionary time; $D$, divergence; r, evolutionary rate) (Yi et al., 2004). LTR pairwise divergence was estimated by the Kimura two-parameter method (Kimura, 1980).

**Nucleotide sequence accession numbers.** Accession numbers for reference ERV sequences in this study were: AY364319 (chromosome A2 endogenous feline leukemia virus, enFeLV-GGAG), AY364318 (chromosome A1 endogenous feline leukemia virus, enFeLV-AGTT), AF052723 (feline leukemia virus strain Rickard subgroup A, FeLV-A), X05470 (baboon endogenous virus, BaEV), EU030001 (rhabdomyosarcoma virus 114, RD-114), AB674444 (feline endogenous retrovirus, ERV-DC10), HM460338 (*Catus familiaris* endogenous retrovirus gamma 1, CiERV), X99935 (murine leukemia-related virus strain VuEV, VuEV), X99934.1 (murine leukemia-related virus strain TaEV, TaEV), X99925.1 (murine leukemia-related virus strain HaEV), X99929.1 (murine leukemia-related virus strain MieVII, MieVII) and X99927.1 (murine leukemia-related virus strain MeVII, MeVII), which are all γ retroviruses; AF033815 (Mason-Pfizer monkey virus, MPMV), which is a β retrovirus; HERV-L (X89211) and MuERV-L (Y12713), which are spuma-like retroviruses.

**ACKNOWLEDGEMENTS**

This work was supported by Konkuk University.

**REFERENCES**


classification of endogenous retroviruses in the canine genome using degenerative PCR and in-silico data analysis. *Virology* 422, 195–204.


