Genomically Intact Endogenous Feline Leukemia Viruses of Recent Origin

Alfred L. Roca, Jill Pecon-Slattery, and Stephen J. O’Brien

Laboratory of Genomic Diversity, Basic Research Program, SAIC-Frederick, and National Cancer Institute, Frederick, Maryland 21702

We isolated and sequenced two complete endogenous feline leukemia viruses (enFeLVs), designated enFeLV-AGTT and enFeLV-GGAG. In enFeLV-AGTT, the open reading frames are reminiscent of a functioning FeLV genome, and the 5’ and 3’ long terminal repeat sequences are identical. Neither endogenous provirus is genetically fixed in cats but polymorphic, with 8.9 and 15.2% prevalence for enFeLV-AGTT and enFeLV-GGAG, respectively, among a survey of domestic cats. Neither provirus was found in the genomes of related species of the Felis genus, previously shown to harbor exFeLVs. The absence of mutational divergence, polymorphic incidence in cats, and absence in related species suggest that these enFeLVs may have entered the germ line more recently than previously believed, perhaps coincident with domestication, and reopens the question of whether some enFeLVs might be replication competent.

Endogenous feline leukemia virus (enFeLV) sequences are present in the genome of the domestic cat, Felis catus, with an estimated 6 to 12 copies per haploid genome (4, 17, 28, 30, 31). These sequences are homologous to exogenous FeLVs (exFeLVs), which are horizontally transmitted oncogenic retroviruses capable of inducing both proliferative and degenerative diseases (12, 27). Endogenous feline leukemia proviruses are part of the germ line and are transmitted from parent to offspring as integral components of chromosomes (4, 17). enFeLV sequences do not produce infectious virus, and attempts to rescue or induce endogenous virus by cocultivation in cell lines have failed (3, 19).

Restriction enzyme mapping has revealed the presence of large deletions in some enFeLVs, which would render them incapable of producing an exogenous virus (41, 42). Molecular copies of enFeLVs without major deletions were found in DNA transfection studies also to be noninfectious, presumably owing either to alterations in sequences regulating gene expression or to coding sequence mutation (42). Both frameshift and nonsense mutations have been identified in the gag and env regions of full-length enFeLVs (5, 27).

Although they do not produce infections on their own, enFeLV sequences readily recombine with exFeLVs (32, 37, 43). Transmissible exFeLVs lack recombinant enFeLV segments and are classified as subgroup A (12, 15). The two other exFeLV subgroups, B and C, result from recombination between enFeLV segments and exFeLVs (27, 32, 35, 43). Recombinant viruses may exhibit altered biological activity and pathogenicity (13, 33, 37, 39, 47); for example, the recombinant subgroup C viruses have been found to induce aplastic anemia (14). Additionally, segments of enFeLVs are transcribed and translated in lymphoma and other cell lines; a truncated enFeLV envelope protein has been detected that inhibits infection by subgroup B exFeLVs (24). Transcription and translation of enFeLV genes have also been demonstrated in tissues from healthy cats, including lymphoid tissue, raising the prospect of a protective role for enFeLVs in vivo (6, 24). In contrast, a protein derived from an enFeLV env region was found to facilitate infection by a T-cell-tropic exFeLV (1).

Despite their possible role in protecting against infection by exFeLVs, and their established capacity to recombine with exFeLVs to produce new strains, the genomic structure and variation of enFeLVs have not been well characterized. Although sequences of endogenous long terminal repeats (LTRs) (5, 18), env (18), pol (33), and part of gag (5) have been determined, the full sequence of a complete enFeLV has not been reported. We therefore generated a probe from a 7-kb gag-pol-env segment (pKHR2-gpe; see Appendix) of a recombinant subgroup B exFeLV (pKHR2/αHF60) (11, 26) and screened a domestic cat lambda FIX II genomic library (9- to 23-kb insert size; Stratagene) (36). Two previously undescribed full-length enFeLVs, designated enFeLV-AGTT and enFeLV-GGAG (the distinguishing label is the unique 4-bp segment of host DNA duplicated during viral integration), were isolated and sequenced. The proviral genome was 8,695 bp long for enFeLV-AGTT and 8,667 bp long for enFeLV-GGAG. These are longer than the 8,440- to 8,448-bp genomes of the two nonrecombinant exFeLVs whose complete sequences are available (GenBank accession numbers M18247 [10] and AF052723 [8]). They are also longer than the 8.2 kb previously estimated by restriction fragment analysis for a full-length enFeLV (42). The gag, pol, and env regions of the two novel proviruses were closer in sequence to exFeLVs than to exFeLVs. For example, enFeLV-AGTT pol had 98.3% nucleotide sequence identity to endogenous L06140 pol but only 95.4% identity to exFeLV M18247 pol. The sequences of enFeLV-AGTT and enFeLV-GGAG were remarkably similar, differing by only a single substitution in the 1,512-bp gag region and by eight substitutions in the 3,630-bp pol region. The length of the env region was 2,009 bp in enFeLV-AGTT and 2,010 bp in enFeLV-GGAG, with two nucleotide substitutions (including one in the region of overlap between pol and env) and one
The novel endogenous proviral sequences were aligned versus previously characterized enFeLV and exFeLV sequences with CLUSTALX (45), and phylogenetic analyses were implemented in PAUP*4.0b4 (44) with three different methods (neighbor joining [NJ], maximum parsimony [MP], and maximum likelihood [ML]), each of which yielded similar tree topologies. The ML tree for pol is shown in Fig. 1A, and it reflects the closer relationship of the novel proviral segments to endogenous rather than exogenous sequences (also true for gag and env; not shown, see Appendix).

The gag, pol, and env viral coding regions are flanked by noncoding LTRs. Unlike the rest of the proviral enFeLV genome, the U3 region of the LTR is not homologous between enFeLVs and exFeLVs (5). The U3 region forms the 5' end of each proviral LTR (in enFeLV-AGTT, the 568-bp LTR includes a 423-bp U3 region). The U3 region has been used as a straightforward means of distinguishing between exFeLV and enFeLV sequences in hybridization studies (7, 29), and the U3 sequences of enFeLV-AGTT and enFeLV-GGAG readily identified them as enFeLVs and not exFeLVs. Because of the absence of homology, exFeLV LTRs were not included in the LTR phylogenetic analyses, which revealed that enFeLV LTRs cluster into two groups (Fig. 1B; Appendix). The U3 regions of the LTRs of exFeLVs are too dissimilar for alignment with those of endogenous LTRs; thus, these were excluded from this analysis. ML tree — In likelihood = 1,004.08144. Subdivision of endogenous LTRs into two groups was also supported by NJ and MP analyses, which generated the same tree topology (best tree found by MP: length = 102, CI = 0.941, RI = 0.969).

FIG. 1. Phylogenetic analyses of proviral regions from enFeLV-AGTT and enFeLV-GGAG and sequences in the GenBank database. ML trees are depicted, drawn by midpoint rooting, with bootstrap support (100 iterations) listed above branches for nodes supported by all three methods: NJ (left), MP (middle), and ML (right). The novel sequences enFeLV-AGTT and -GGAG are compared to previously published sequences labeled with their GenBank accession numbers. FeLVA, exFeLV subgroup A. (A) Analyses of sequences from the pol viral region demonstrate that enFeLV-AGTT and enFeLV-GGAG are more closely related to enFeLV than to exFeLV sequences. The full-length sequence was used to generate the tree for pol (3,633 bp). The score (−ln likelihood) of the best ML tree was 6,421.51427; the same tree topology was produced by NJ and MP (best tree found by MP: length = 277, consistency index [CI] = 0.968, retention index [RI] = 0.941). (B) ML tree for the full-length (570-bp) proviral LTRs of enFeLVs reveals their subdivision into two sets of sequences, designated groups I and II. The U3 regions of the LTRs of exFeLVs are too dissimilar for alignment with those of endogenous LTRs; thus, these were excluded from this analysis. ML tree − In likelihood = 1,004.08144. Subdivision of endogenous LTRs into two groups was also supported by NJ and MP analyses, which generated the same tree topology (best tree found by MP: length = 102, CI = 0.941, RI = 0.969).
framen shift mutation in the coding region for the gp70 protein (arrow in Fig. 2) (10). This site (residue 200) contains a succession of nine cytosines in the undisrupted enFeLV-AGTT coding sequence. In enFeLV-GGAG, a 10th cytosine is present in the poly(C) region, presumably resulting after strand slippage during DNA replication. Another mutation disrupts the putative start codon for env in enFeLV-GGAG. A polymorphic distribution (i.e., integrated virus versus empty chromosomal sites) of enFeLVs has been suggested by Southern blotting (17). We determined the distribution among cats of enFeLV-AGTT and enFeLV-GGAG by using one PCR primer based on the genomic DNA flanks unique to each individual enFeLV, with a second primer based on the proviral sequence (Fig. 3 and Appendix). Cats from different genetic backgrounds and geographic origins, including individuals from nine recognized breeds, were screened for the presence of enFeLV-AGTT and enFeLV-GGAG (Table 1 and Appendix). Among the 79 domestic cats screened, enFeLV-AGTT was present in 7 individuals (8.9% of the cats, 4.4% of the domestic cat lineage, which are known to carry enFeLVs (Appendix) (3, 21, 23). The presence of enFeLVs in only these domestic cat lineage before the lineage radiated (2, 3, 17, 27), i.e., millions of years ago (23). Neither enFeLV-AGTT nor enFeLV-GGAG was found to be present in any of the wild cats tested with multiple primer pairs, although both were discovered in the original cat used to construct the genomic library.

We also screened for the presence of enFeLV-AGTT and enFeLV-GGAG in individuals from wild Felis species of the domestic cat lineage, which are known to carry enFeLVs (Appendix) (3, 21, 23). The presence of enFeLVs in only these species of felids has suggested that enFeLVs entered the germ line of a common ancestor of the domestic cat lineage before the lineage radiated (2, 3, 17, 27), i.e., millions of years ago (23). Neither enFeLV-AGTT nor enFeLV-GGAG was found to be present in any of the wild cats tested with multiple primer pairs, although primers spanning the proviral integration site readily amplified both in Felis species and in more distantly three Turkish Van cats had enFeLV-AGTT. Three nonbreed cats, including a feral cat from Australia, were also found to have enFeLV-AGTT. All seven of the cats were heterozygous for the presence of enFeLV-AGTT, since a PCR designed to span the proviral integration site was also successful, indicating that enFeLV-AGTT was present in only one of the two sister chromosomes. Because enFeLV-AGTT is found in a small minority of domestic cats, it may have been absent from the cell lines used in induction studies, in which FeLVs could not be induced from enFeLVs (3, 19, 42). While this result was attributed to deletions or mutations in enFeLVs, induction of FeLV be induced from enFeLVs (3, 19, 42). While this result was attributed to deletions or mutations in enFeLVs, induction studies have not been attempted with cell lines screened for the presence of the undisrupted enFeLV-AGTT provirus, which reopens the question of whether some enFeLVs might be replication competent.

Among the 79 domestic cats (Table 1 and Appendix), enFeLV-GGAG was present in 12 (15.2% of the cats, 8.2% of the chromosomes), including 4 of the 6 Persian cats screened and 1 of 2 Siamese cats (Table 1). The other seven enFeLV-GGAG-positive cats were nonbreed cats or of unknown lineage. One cat, Fca 215, was homozygous for the presence of the provirus, as repeated attempts to amplify an unintegrated copy present on both chromosome homologues; +/−, copy present on one of two homologues; −/−, no copies present.

### Table 1. Domestic cats with enFeLV-AGTT or -GGAG

<table>
<thead>
<tr>
<th>Reference no.</th>
<th>enFeLV-AGTT</th>
<th>enFeLV-GGAG</th>
<th>Breed or locale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fca 138</td>
<td>+/−</td>
<td>−/−</td>
<td>Egyptian Mau</td>
</tr>
<tr>
<td>Fca 146</td>
<td>+/−</td>
<td>−/−</td>
<td>Egyptian Mau</td>
</tr>
<tr>
<td>Fca 569</td>
<td>−/−</td>
<td>+/−</td>
<td>Turkish Van</td>
</tr>
<tr>
<td>Fca 828</td>
<td>+/−</td>
<td>−/−</td>
<td>Turkish Van</td>
</tr>
<tr>
<td>Fca 171</td>
<td>+/−</td>
<td>−/−</td>
<td>Australia, feral</td>
</tr>
<tr>
<td>Fca 84</td>
<td>+/−</td>
<td>−/−</td>
<td>United States</td>
</tr>
<tr>
<td>Fca 165</td>
<td>+/−</td>
<td>−/−</td>
<td>United States</td>
</tr>
<tr>
<td>Fca 215</td>
<td>−/−</td>
<td>+/−</td>
<td>Persian</td>
</tr>
<tr>
<td>Fca 997</td>
<td>−/−</td>
<td>+/−</td>
<td>Persian</td>
</tr>
<tr>
<td>Fca 1082</td>
<td>−/−</td>
<td>+/−</td>
<td>Persian</td>
</tr>
<tr>
<td>Fca 1093</td>
<td>−/−</td>
<td>+/−</td>
<td>Persian</td>
</tr>
<tr>
<td>Fca 761</td>
<td>−/−</td>
<td>+/−</td>
<td>Siamese</td>
</tr>
<tr>
<td>Fca BW7</td>
<td>−/−</td>
<td>+/−</td>
<td>Britain</td>
</tr>
<tr>
<td>Fca 38</td>
<td>−/−</td>
<td>+/−</td>
<td>United States</td>
</tr>
<tr>
<td>Fca 42</td>
<td>−/−</td>
<td>+/−</td>
<td>United States</td>
</tr>
<tr>
<td>Fca 127</td>
<td>−/−</td>
<td>+/−</td>
<td>United States</td>
</tr>
<tr>
<td>Fca 129</td>
<td>−/−</td>
<td>+/−</td>
<td>United States</td>
</tr>
<tr>
<td>Fca FAS11</td>
<td>−/−</td>
<td>+/−</td>
<td>United States</td>
</tr>
<tr>
<td>Fca FAS5</td>
<td>−/−</td>
<td>+/−</td>
<td>United States</td>
</tr>
</tbody>
</table>

* +/−, copy present on both chromosome homologues; +/−, copy present on one of two homologues; −/−, no copies present.
related Felidae species (Appendix). The absence of enFeLV-AGTT among wild cats and lack of fixation among domestic cats raise the possibility that integration of enFeLV-AGTT occurred subsequent to the domestication of cats. The viruses that produced enFeLVs were not thought to have persisted except as molecular “fossils” in the genome of the ancestor of the domestic cat lineage (27). The enFeLV-AGTT provirus suggests more recent persistence of these FeLVs among domestic cats or related Felis species. Since enFeLVs may derive from rodent viruses (2, 3), the possibility that viruses emerged from rodents on multiple occasions also cannot be excluded.

Retroviral 5′ and 3′ LTRs are identical in sequence at the time of integration, although random mutation would cause proviral 5′ and 3′ LTR sequences to gradually drift apart after incorporation into the host germ line (16). The 5′ and 3′ LTRs of enFeLV-GGAG were different from each other at two nucleotide sites, but the 5′ and 3′ LTR sequences of enFeLV-AGTT are identical, which suggests that integration of the latter provirus occurred relatively recently in the evolutionary history of cats. The substitution rate for enFeLV-AGTT versus human (HERV-K113; 970-bp combined LTRs) is 1.2% per 10^6 years (20), then we consider a divergence rate estimate for noncoding regions of enFeLV-GGAG proviruses to begin within 170,000 to 385,000 years after the radiation of species within the domestic cat lineage. Since the presence of enFeLVs at the same genomic location in two individuals is an indication of common ancestry, enFeLVs may prove useful as genetic markers for establishing relationships among individuals, lineages, and species within the genus Felis.

**Nucleotide sequence accession numbers.** The sequences of the novel enFeLVs described here have been deposited in the GenBank database (accession numbers AY364318 and AY364319).

**APPENDIX**

For each of the previously published sequences used in phylogenetic analysis, the accession number is included as part of the sequence. Felids with neither enFeLV-AGTT nor enFeLV-GGAG present were as follows: domestic cat, Felis catus, by breed or locale, Abyssinian breed, Fca 567, 618, and 641; American Shorthair, Fca 326, 327, 329, and 391; Birman, Fca 620 and 626; Burmese breed, Fca 9, 364, 366, 376, 379, 381, 382, 384, 385, 386, 387, 388, 389, and 390; Havana Brown, Fca 792; Japanese Bobtail, Fca 599, 600, and 603; Persian breed, Fca 1061 and 1067; Russian Blue, Fca 1094 and 1095; Siamese breed, Fca 559; Turkish Van, Fca 583; Argentina, Fca 157; Australia (ferral), Fca 168 and 169; Britain, Fca GWK, GW1, and TB4; Costa Rica, Fca 150; Russia, Fca 140; United States, Fca 12, 17, 18, 21, 23, 24, 39, 52, 122, 123, 132, 133, 136, 186, 223, 264, 265, and FAS13. Wild species of the domestic cat lineage: Felis bioti, Chinese mountain cat, Fbi 2; Felis chaus, jungle cat, Fch 1, 2, 4, and 5; Felis lybica, African wild cat, Fli 3; Felis margarita, sand cat, Fma 5, 8, 10, 11, and 13; Felis nigripes, black-footed cat, Fni 3, 4, 5, 6, and 14; Felis silvestris, European wild cat, Fsi 1, 6, 7, 9, 13, 18, 21, and 25. Other wild felid species: Herpaillurus yagouaroundi, jaguarundi, Hya 12; Leopards wiedii, mar- gay, Lwi 19 and 70; Lynx pardinum, Iberian lynx, Lpa 11; Lynx rufus, bobcat, Lru 38 and 43; Otocologus manul, Pallas cat, Oma 3, 4, 5, 10, 14, and 15; Panthera leo, lion, Ple 7; Panthera uncia, snow leopard, Pun 13; Puma concolor, puma, Pco 333.

The PCR primers used to generate the pKHR2-gpe DNA for library screening were GA-GAG-F1 (ATGGGGCCAAACTAATACTACC) and GA-ENV-R1 (TGGTGGTGGAGATCATATTG). The long PCR used to isolate each LTR on a separate DNA fragment used one primer based on the left phage arm (FIXII-IA; GCGCCGGCGGC TCTAATTACA) or the right phage arm (FIXII-RA; GCGCCGGCC GAGCTCAATTAACC) and a second primer based on the enFeLV pol sequence, in either the forward (POL-F8XL; ACCRAOGGAAA ACTATAAGGCTGTA) or the reverse (POL-R8XL; GCCCCAGCCA GAGAAGGTTCTAT) direction. PCR screening for the presence of enFeLV-AGTT was done with primers 6FL5-F1 (CCTTGATTAGA AGGTAAGGT) and LTR-R4 (CTCAGCAAAAGGACTTGCG), primers LTR-F8 (AAACAGGATATCTGTGGTCA) and 6FL3-R4 (ATTCCTTAATACACTTGAT), primers 6-5F1L (CCCRGCGT GTAGGAAAT) and LTR-R2L (CRGGTGGCTGGACACAGA TA), or primers LTR-F2L (GCCGAACTTTGTGCTAG), and 6-3R1L (TGAAAATCGAAGAAAGACGAGG). For absence of enFeLV-AGTT, the primer combinations used were 6FL5-F1 (CTTC AGTGCATAACAAG) and 6FL3-R3 (TTGCAATTTTTAGAGA TAGTCAA), 6FL5-F3 and 6FL3-R4 (ATTCCTTAATACACTTGAT), 6FL5-F4 (TTCCTGATGGGGACAGT) and 6FL3-R3, and 6-5F1L and 6-3R1L. For the presence of enFeLV-GGAG, the primers used were LTR-F8 and 16FL3-R4 (CAACTCCTTGTGACATCG), 16FL5-F3 (TGGCAAGAAGCTGGTGA) and LTR-R4, 16-5F1L (TTCAGAGAGAACCGTGA) and LTR-R2L, and LTR-F2L and 16-3R1L (AAGGAGACCCTAAGGAGG). The primers used to test for the absence of the enFeLV-GGAG were 6FL5-F1 (AAAC AGAAGACCTACAAG) and 16FL3-R4, 16FL5-F3 and 16FL3-R4, 16FL5-F4 (TTCCTCACCTGTCCT) and 16FL3-R3 (CTCTTACCTGTCCT).
sequence consisting of an alignment of 322 characters at the 5’ end of the 
region (of which 46 were parsimony informative in the MP 
analysis). Full-length sequences were used for 
env (3,633 characters, 153 parsimony informative), and for LTRs (570 characters, 82 parsimony 
informative). MP analyses treated gaps as a 
region (of which 46 were parsimony informative in the MP 
analysis). Full-length sequences were used for 
env (3,633 characters, 153 parsimony informative), and for LTRs (570 characters, 82 parsimony 
informative). MP analyses treated gaps as a 


with variant viruses. Leukemia 6:1538–1545.


envelope genes of feline leukemia virus-related 
sequences in cats and attemps to identify enoxogenous viral sequences in tissues of virus-negative 


cat cell clone with properties distinct from previously described feline type C 


cells: evidence for a role in natural resistance to subgroup B feline leukemia 


mia virus proviruses from a virus-productive human cell line. J. Virol. 38: 
688–703.

Retrovirus biology and human disease. Dekker, New York, N.Y.

28. Nimman, H. L., M. Akhavi, M. B. Gardner, J. R. Stephenson, and P. Roy-
Burman. 1980. Differential expression of two distinct endogenous retrovirus 

envelope virus genome is not endogenous in cat cells. Int. J. Cancer 22:70–78.

cell DNA. Intervirology 12:253–260.

Olsen. 1976. FeLV-related sequences in DNA from a FeLV-free cat colony. 
Intervirology 6:798–801.


33. Pandey, R., A. K. Ghosh, D. V. Kumar, B. A. Bachman, D. Shibata, and P. Roy-
Burman. 1991. Recombination between feline leukemia virus subgroup 
B or C and endogenous env elements alters the in vitro biological activities of the 


and host range determinants of the feline aplastic anemia retrovirus. Proc. 

36. Roca, A. L., C. Godson, D. R. Weaver, and S. M. Reppert. 1996. Structure, 
characterization, and expression of the gene encoding the mouse Mel1a 


feline leukemia virus genes detected in naturally occurring feline lympho-

40. Shih, A. E. Coutavas, and M. G. Rush. 1991. Evolutionary implications of 
