

## Evolutionary dynamics of endogenous feline leukemia virus proliferation among species of the domestic cat lineage

Sagi Polani <sup>a,1</sup>, Alfred L. Roca <sup>b,\*</sup>, Bryan B. Rosensteel <sup>c</sup>, Sergios-Orestis Kolokotronis <sup>d</sup>, Gila Kahila Bar-Gal <sup>a,\*</sup>

<sup>a</sup> The Koret School of Veterinary Medicine, The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot, 76100, Israel

<sup>b</sup> Department of Animal Sciences and Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

<sup>c</sup> University of Maryland, Baltimore, MD 21201, USA

<sup>d</sup> Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY 10024, USA

### ARTICLE INFO

#### Article history:

Received 31 March 2010

Returned to author for revision 22 April 2010

Accepted 3 June 2010

Available online 7 July 2010

#### Keywords:

Endogenous retroviruses

Oncoretroviruses

Domestic cat

African wildcat

European wildcat

Sand cat

Jungle cat

Black-footed cat

### ABSTRACT

Endogenous feline leukemia viruses (enFeLVs) occur in the germ lines of the domestic cat and related wild species (genus *Felis*). We sequenced the long terminal repeats and part of the *env* region of enFeLVs in domestic cats and five wild species. A total of 305 enFeLV sequences were generated across 17 individuals, demonstrating considerable diversity within two major clades. Distinct proliferations of enFeLVs occurred before and after the black-footed cat diverged from the other species. Diversity of enFeLVs was limited for the sand cat and jungle cat suggesting that proliferation of enFeLVs occurred within these species after they diverged. Relationships among enFeLVs were congruent with host species relationships except for the jungle cat, which carried only enFeLVs from a lineage that recently invaded the germline (enFeLV-AGTT). Comparison of wildcat and domestic cat enFeLVs indicated that a distinctive germ line invasion of enFeLVs has not occurred since the cat was domesticated.

© 2010 Elsevier Inc. All rights reserved.

### Introduction

Endogenous feline leukemia viruses (enFeLVs) are proviral elements within the genome of the domestic cat and closely related wild species of felids (Coffin, 2004; Roca et al., 2004; Weiss, 2006). By themselves, enFeLVs do not code for infectious virus, due to mutations present in the proviral genome (Berry et al., 1988; Kumar et al., 1989; Mullins and Hoover, 1990; Stewart et al., 1986). However, recombination occurs between inherited enFeLV and exogenous feline leukemia virus (exFeLV) following infection, which can alter the genetic makeup of the exogenous virus, giving rise to new infectious agents that may result in severe pathogenesis (Flint et al., 2004; McDougall et al., 1994; Mullins and Hoover, 1990; Overbaugh et al., 1988; Rezanka et al., 1992; Roy-Burman, 1995; Stewart et al., 1986; Tzavaras et al., 1990; Weiss, 2006).

The number of copies of enFeLV per haploid genome of the domestic cat has been estimated as 6–12 (Benveniste et al., 1975; Koshy et al., 1980; Niman et al., 1980) while fluorescent in situ hybridization detected 9–16 distinct autosomal loci of enFeLVs per domestic cat examined (Roca et al., 2005). Within the cat family Felidae, enFeLVs are present only among species that comprise the domestic cat lineage, i.e., members of the genus *Felis*. In addition to the domestic cat (*Felis catus*), wild species within the lineage include: *Felis lybica* (designated “Fli”, African wildcat); *Felis silvestris* (designated “Fsi”, European wildcat); *Felis bieti* (designated “Fbi”, Chinese desert cat); *Felis margarita* (designated “Fma”, sand cat); *Felis nigripes* (designated “Fni”, black-footed cat); and *Felis chaus* (designated “Fch”, jungle cat) (Driscoll et al., 2007; Johnson et al., 2006). However, enFeLVs are not found in the genomes of other lineages within the Felidae (Benveniste et al., 1975; Johnson et al., 2006). Thus, primary colonization of the germ line by enFeLVs is believed to have occurred in an ancestor of the domestic cat lineage, before the divergence of any of the branches leading to the extant species, but after the initial divergence of the domestic cat lineage from the rest of the cat family (Benveniste et al., 1975; Johnson et al., 2006). More recently, phylogenetic studies of the LTR sequences of enFeLVs in domestic cats have demonstrated that enFeLV proviruses in that species fall into two divergent clades (Pontius et al., 2007; Roca et al., 2004). This suggests that two separate invasions of enFeLVs may have occurred in the ancestry of domestic cats, while at least some of the proviruses

\* Corresponding authors. A.L. Roca is to be contacted at 432 ASL, 1207 West Gregory Drive, Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. Fax: +1 217 333 8286. G.K. Bar-Gal, The Koret School of Veterinary Medicine, The Hebrew University of Jerusalem, Rehovot, 76100, Israel. Fax: +972 8 946 7940.

E-mail addresses: [sagi.polani@gmail.com](mailto:sagi.polani@gmail.com) (S. Polani), [roca@illinois.edu](mailto:roca@illinois.edu) (A.L. Roca), [bryanr1@umbc.edu](mailto:bryanr1@umbc.edu) (B.B. Rosensteel), [koloko@amnh.org](mailto:koloko@amnh.org) (S.-O. Kolokotronis), [bargal@agri.huji.ac.il](mailto:bargal@agri.huji.ac.il) (G.K. Bar-Gal).

<sup>1</sup> Contributed equally.

within one of the enFeLV groups have an evolutionarily recent origin (Roca et al., 2004).

The discovery that enFeLVs may have invaded the germ line of domestic cat ancestors at more than one interval during evolutionary history, and that some enFeLVs may be of evolutionarily recent origins, led us to formulate and test a number of hypotheses regarding the invasion of the ancestral germ line of *Felis* by enFeLVs. First, since two distinct clades of enFeLV were detected in the domestic cat, it seemed possible that one of the two clades may have a more ancient evolutionary origin than the other clade. This would be supported if one of the clades of enFeLV were found present in more recently diverged wild species of *Felis* than the other enFeLV clade. The presence or absence of the two different types of enFeLV across species within the domestic cat lineage would help to identify the evolutionary intervals during which different FeLVs invaded the ancestral germ line. Second, we hypothesized that sequencing enFeLVs from wild species within the domestic cat lineage might reveal additional subgroups of enFeLVs not present in the domestic cat genome, and distinctive from the two subgroups of enFeLV previously identified in the domestic cat. The detection of novel clades of enFeLV would contribute to understanding the evolutionary history of the domestic cat lineage. Third, whether or not additional highly distinctive clades of enFeLV were uncovered, we also hypothesized that there may be diversity within the two previously identified enFeLV clades, and that subclades may be present within the major enFeLV clades that correspond only to one or a few of the species within the domestic cat lineage. We looked for evidence that specific types of enFeLVs had affected only one of the species in the lineage, either through a separate invasion of its germ line by exogenous enFeLVs, or through proliferation or homogenization of enFeLVs already present in the genome after divergence of the species. Fourth, we wished to test the hypothesis that the process of domestication may have affected enFeLVs in domestic cats. Exogenous FeLV infections are common among large groups of cats kept at high densities (Lee et al., 2002), and domestic cats are often kept at much higher densities than would be typical for wildcats. The higher density at which domestic cats are kept may have influenced their patterns of enFeLV when compared to their wildcat progenitors.

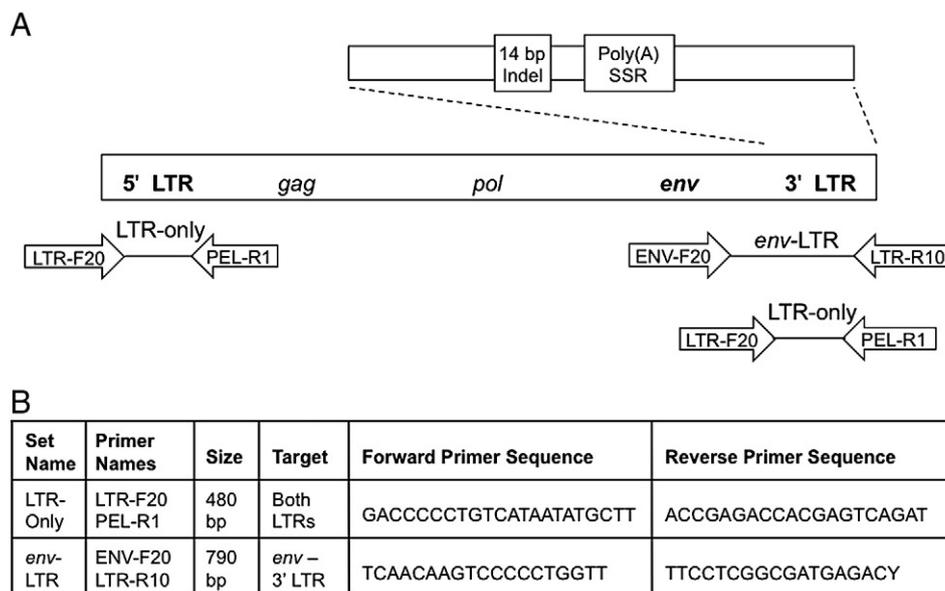
To test these four hypotheses, PCR and sequencing were conducted for enFeLVs using samples from domestic and all wild species of felids within the domestic cat lineage. To maximize the diversity of sequences collected, two different primer pairs were used, designed to amplify overlapping regions of the proviral genome. Amplification, cloning, sequencing and analysis of enFeLV proviruses revealed the evolutionary history of enFeLVs in the domestic cat and related wild species.

## Results

### Sequence diversity of enFeLVs

One pair of conserved primers (designated “env-LTR”) was used to amplify and sequence a region of enFeLV including the *env* and 3' LTR, avoiding solo LTRs that lack any coding sequence and that are known to outnumber enFeLVs with coding regions (Casey et al., 1981). The other pair of conserved primers (designated “LTR-only”) amplified both 5' and 3' LTRs (Fig. 1), including solo LTRs. PCR was attempted on 25 individuals across all seven species of the domestic cat lineage (Table 1). However, due to poor DNA quality, eight samples (Fca-215, Fli-1, Fbi-3, Fbi-5, Fbi-6, Fma-4, Fma-11 and Fch-2) could not be successfully amplified using either primer set (Table 1), even after several amplification attempts. For *Felis bieti*, the rare Chinese desert cat, the DNA was of poor quality for all three available samples (Table 1), and this species was therefore excluded from the study. For each of the other six species of *Felis*, PCR was successful for up to six individuals from each species (Table 1). Sequences were successfully generated for a total of 17 individuals (Table 1); for one of these individuals, Fli-3, only the env-LTR primer set was successfully cloned; while for Fsi-5 just the LTR-only primer set was successful.

A total of 305 sequences were generated for enFeLV clones, including 148 for the env-LTR region and 157 for the LTR-only region (Table 1). Among the clones, 18 within-individual sequences produced for env-LTR were identical, and 22 within-individual sequences produced for LTR-only were identical. These duplicate sequences within the same individual were removed from most analyses (see Materials and methods for a listing of duplicates). Two sequences were excluded from the env-LTR dataset due to short length



**Fig. 1.** Endogenous feline leukemia proviral structure showing sequenced regions, variable sequence motifs and primer sequences. (A) Endogenous FeLV proviral genome structure showing the positions of amplicons (env-LTR and LTR-only) targeted by two sets of primers, the names and locations of the primers used for each set, and the locations of a variable-length poly(A) mononucleotide simple sequence repeat (SSR) and of a 14-bp indel within both long terminal repeats (LTRs). These would be present on both LTRs but are only illustrated for the 3' LTR. (B) Forward and reverse oligonucleotide primer sequences used for PCR and sequencing of enFeLVs.

**Table 1**  
Species of the domestic cat lineage: sample information, and sequences generated.

| Species                                    | Sample name | Geographic origin                     | env-LTR primers |                    |       | LTR-only primers |                    |       |
|--|-------------|---------------------------------------|-----------------|--------------------|-------|------------------|--------------------|-------|
|  |             |                                       | No. of clones   | Sequence diversity | S.E.  | No. of clones    | Sequence diversity | S.E.  |
| <i>Felis catus</i> (domestic cat)          | Fca-9       | Burmese breed, Southeast Asia         | 10              | 0.012              | 0.005 | 10               | 0.025              | 0.005 |
|  | Fca-127     | Non-breed, Washington State, USA      | 10              | 0.015              | 0.006 | 9                | 0.028              | 0.005 |
|  | Fca-146     | Egyptian Mau, Egypt                   | 10              | 0.024              | 0.006 | 7                | 0.026              | 0.005 |
|  | Fca-215     | Persian breed, Iran                   | N.A.            | N.A.               | N.A.  | N.A.             | N.A.               | N.A.  |
| <i>Felis lybica</i> (African wildcat)      | Fli-1       | Unknown                               | N.A.            | N.A.               | N.A.  | N.A.             | N.A.               | N.A.  |
|  | Fli-3       | Tajikistan                            | 9               | 0.002              | 0.001 | N.A.             | N.A.               | N.A.  |
| <i>Felis silvestris</i> (European wildcat) | Fsi-1       | Azerbaijan                            | 8               | 0.007              | 0.002 | 10               | 0.034              | 0.006 |
|  | Fsi-5       | <i>F. s. tristrami</i> , Saudi Arabia | N.A.            | N.A.               | N.A.  | 10               | 0.008              | 0.003 |
|  | Fsi-7       | <i>F. s. gordonii</i> , Saudi Arabia  | 10              | 0.012              | 0.004 | 12               | 0.014              | 0.005 |
|  | Fsi-9       | <i>F. s. gordonii</i> , Saudi Arabia  | 11              | 0.010              | 0.004 | 11               | 0.027              | 0.005 |
|  | Fsi-13      | <i>F. s. ornata</i> , India           | 9               | 0.009              | 0.002 | 12               | 0.008              | 0.003 |
|  | Fsi-24      | <i>F. s. caudata</i> , Turkmenistan   | 9               | 0.014              | 0.003 | 9                | 0.025              | 0.006 |
| <i>Felis bieti</i> (Chinese desert cat)    | Fbi-3       | Gansu, China                          | N.A.            | N.A.               | N.A.  | N.A.             | N.A.               | N.A.  |
|  | Fbi-5       | Qinghai, China                        | N.A.            | N.A.               | N.A.  | N.A.             | N.A.               | N.A.  |
|  | Fbi-6       | Gansu, China                          | N.A.            | N.A.               | N.A.  | N.A.             | N.A.               | N.A.  |
| <i>Felis margarita</i> (sand cat)          | Fma-4       | Unknown                               | N.A.            | N.A.               | N.A.  | N.A.             | N.A.               | N.A.  |
|  | Fma-8       | Unknown                               | 11              | 0.002              | 0.001 | 8                | 0.014              | 0.003 |
|  | Fma-10      | <i>F. m. scheffeli</i> , Pakistan     | 6               | 0.004              | 0.001 | 9                | 0.013              | 0.003 |
|  | Fma-11      | <i>F. m. scheffeli</i> , Pakistan     | N.A.            | N.A.               | N.A.  | N.A.             | N.A.               | N.A.  |
| <i>Felis nigripes</i> (black-footed cat)   | Fni-6       | South Africa                          | 10              | 0.023              | 0.004 | 12               | 0.041              | 0.007 |
|  | Fni-14      | South Africa                          | 11              | 0.025              | 0.004 | 7                | 0.023              | 0.005 |
|  | Fni-15      | Unknown                               | 9               | 0.022              | 0.004 | 12               | 0.018              | 0.003 |
| <i>Felis chaus</i> (jungle cat)            | Fch-1       | Unknown                               | 7               | 0.002              | 0.001 | 10               | 0.001              | 0.001 |
|  | Fch-2       | Unknown                               | N.A.            | N.A.               | N.A.  | N.A.             | N.A.               | N.A.  |
|  | Fch-11      | Unknown                               | 8               | 0.003              | 0.001 | 9                | 0.001              | 0.001 |
| Total clones sequenced:                    |             |                                       | 148             |                    |       | 157              |                    |       |

## Notes:

For domestic cats, geographic origin refers to the region where a breed is believed to have been developed.

N.A., not applicable, refers to samples from which no successful sequences were obtained (e.g., due to poor DNA quality).

S.E., standard error.

relative to the rest. Thus, subsequent analyses of enFeLVs included a total of 128 sequences of the env-LTR region and 135 sequences of the LTR-only region. For analyses that combined the env-LTR and LTR-only datasets, 12 sequences identical between the two datasets were also removed, yielding a total of 251 sequences.

Alignments of the two sets of sequences indicated that there was considerable diversity among species, among individuals within a species, and within each individual. This diversity included point mutations, indels, and a variable-length poly(A) mononucleotide simple sequence repeat (SSR) (Fig. 1). The poly(A) repeat was observed within all LTR sequences for all individuals but varied in length from a minimum of 3 to maximum of 52 adenine nucleotides that in some sequences were interrupted (mainly by guanine and cytosine). Such complexity is common among microsatellites (Culver et al., 2001). The poly(A) SSR corresponded to positions 874–917 (5' LTR) and 9001–9044 (3' LTR) in the enFeLV-AGTT full-length proviral “reference” sequence (Genbank accession AY364318). Additionally, a 14 bp LTR indel was detected (positions 754–774 and 8881–8901 in enFeLV-AGTT) that had identical 5 bp flanks on either side (5'-AAGATAGCCACCTGGCCCTAAGAT-3') with one copy of the flank present in enFeLVs lacking the indel. The insertion was present in a total of 46 of the sequences (14 of the env-LTR sequences and 32 of the LTR-only), including those of four species: *F. nigripes*, *F. silvestris*, *F. lybica* and *F. catus*. This insertion was not detected among the sequences for *F. chaus* or *F. margarita* although the apparent absence of the insert from these two other species may reflect an abundance of enFeLVs lacking the insertion in *F. chaus* and *F. margarita* (see below), rather than a true absence of the insertion from any enFeLVs in their genomes. Using a BLAST query, the insertion was also present in the sequence of three previously published domestic cat enFeLVs (GenBank accession numbers: AY364318.1, AY364319.1, AY364322.1) (Roca et al., 2004).

## Phylogenetic and recombination analyses of enFeLVs

Phylogenetic analysis was performed using the 128 final sequences (after removal of within-individual duplicates) generated using primer pair env-LTR, which extended from the 3' end of env to the 3' LTR of proviral enFeLV. The analysis included previously published enFeLV sequences from *F. catus* that are available in GenBank and that overlapped completely the relevant region of enFeLV (GenBank accessions AY364318.1; AY364319.1) (Roca et al., 2004). The variable-length poly(A) SSR was excluded from the analyses, since it is a microsatellite for which homologous nucleotide positions cannot be established. Phylogenetic analyses of the resulting 722-nt env-LTR alignment revealed two major clades as shown in Fig. 2. BLAST queries allowed the two major clades to be identified as corresponding to those previously identified only in the domestic cat and designated enFeLV Groups I and II (Pontius et al., 2007; Roca et al., 2004). All but one of the sequences forming Group I contained the 14-nt LTR insertion, however, this insertion was also present in three enFeLV sequences of *F. nigripes* in Group II (Fig. 2). We therefore considered the possibility of recombination affecting the sequences.

The reconstruction of the ancestral state at each node of the 14-bp indel examined the possibility of independent deletion events in Group I and II. Parsimony reconstruction showed that the deletion was an original feature of the most recent common ancestor of Group I, but not of any Group II clade containing sequences with the deletion. The likelihood reconstruction method yielded a 0.03642338 symmetrical rate of state change between insertion and deletion. The last common ancestor of the large subclade of the *nigripes*-like groups containing Fni-6\_3\_10 was estimated to have been nine times more likely to have the insertion (proportional likelihood = 0.89657) rather than the deletion (proportional likelihood = 0.10342917). For the other two more phylogenetically distant *F. nigripes* enFeLVs, the



proportional likelihood for deletion to be the ancestral state dropped to  $4.87 \times 10^{-5}$  and  $4.23 \times 10^{-9}$  for Fni-6\_3\_5 and Fni-6\_3\_12, respectively. This could be suggestive of recombination or gene conversion (Hughes and Coffin, 2005) dispersing the 14-bp deletion among *F. nigripes* enFeLVs outside Group I in the *nigripes*-like groups. Nonetheless, using the software RDP 3b41 (Martin et al., 2005) we were unable to find statistically significant evidence for the existence of breakpoints in the *env*, 39-nt spacer, and 3' LTR segments of the *env*-LTR dataset. Furthermore, when the highly diverse poly(A) SSR was compared to the *env*-LTR phylogeny, similar poly(A) SSRs tended to be found among enFeLVs that clustered together. Thus the *env*-LTR sequences did not appear to have been greatly affected by recombination. Thus the *env*-LTR phylogeny likely reflects actual patterns present among enFeLVs in the domestic cat lineage, given the lack of recombination and the large number of sequences, individuals and species analyzed for the dataset.

Using the *env*-LTR primers, Group I sequences were detected primarily in *F. nigripes* with copies also detected in *F. catus* and *F. silvestris* (Fig. 2). Group II sequences were detected in all of the species. Across the tree, clustering of sequences by taxon was evident for some species. Within Group II, a number of subdivisions were apparent. A large number of sequences fell into a subclade that was designated the “enFeLV-AGTT-clade” (Fig. 2). The designation and the basal limit of the clade were established by the position of enFeLV-AGTT, a provirus previously discovered in the domestic cat that appears to be genomically intact (i.e. no disruption of the reading frames) and of recent origin (Roca et al., 2004). The two LTRs of domestic cat enFeLV-AGTT are identical in sequence, suggesting that relatively little time has elapsed since integration (Roca et al., 2004).

The AGTT-clade shows minimal phylogenetic structure with nearly null internode branch lengths and together with the related “AGTT-like groups” exhibited a high degree of clustering for host-specific enFeLV sequences (Fig. 2); namely the jungle cat, *F. chaus* (see above), and the sand cat, *F. margarita*, in AGTT-clade and AGTT-like groups, respectively (Fig. 2). Interestingly, although the jungle cat, *F. chaus*, has one of the two most basal positions within the domestic cat lineage (Johnson et al., 2006), for this dataset all of the enFeLVs for *F. chaus* fell within the AGTT-clade, and none fell within more basal clades of enFeLV (Fig. 2).

Sequences from the African black-footed cat, *Felis nigripes*, were diverse and distributed across both major groups (I and II). *F. nigripes* sequences present within Group II were at a position basal to that of the AGTT-clade and AGTT-like groups. These basal sequences within Group II were therefore designated the “*nigripes*-like groups” (Fig. 2). A number of the subclades within the *nigripes*-like groups included only enFeLV sequences generated from *F. nigripes*. For these subclades, and for the *margarita*-clade (a clade consisting of only *F. margarita* enFeLVs, within the AGTT-like groups) (Fig. 2), the apparent monophyly of the subclades could have been due to a species-specific genomic invasion of the germ line by a particular type of exogenous FeLV. On the other hand, the patterns may only represent differences in the relative abundances of subclade enFeLVs between species, possibly due to secondary proliferation (Bannert and Kurth, 2006) or to homogenization through gene conversion (Hughes and Coffin, 2005) within species. For example, the presence of only *F. margarita* enFeLV sequences within the “*margarita*-clade” may have been due to a separate germ line invasion of enFeLVs that affected only

*F. margarita*. Alternatively, it may be that *margarita*-clade enFeLVs are also present in other species, but at a relative abundance that was too low to be detected by the current survey of enFeLVs. Since the genome of the domestic cat has been sequenced (Pontius et al., 2007), these two alternative hypotheses could be examined using the domestic cat genomic sequence traces. For enFeLV *env*-LTR sequences within the *margarita*-clade, two sites diagnostic of the clade (common to all members of the clade but absent from other enFeLVs in *Felis catus*) were identified. The enFeLV *margarita*-clade diagnostic sites (along with short conserved sequences flanking either side of a diagnostic site) were queried against sequence traces of the domestic cat genome. For the *margarita*-clade as well as for the various clades containing only *F. nigripes*, the individual diagnostic sites were all detected among the domestic cat genomic sequence traces. This suggested that the apparent monophyly of the *margarita*-clade and the various subclades containing only *F. nigripes* enFeLV sequences may be due to different relative abundances of various enFeLVs across species (i.e., there was a relative under-representation of *margarita*-clade sequences in the domestic cat), and are not necessarily due to separate genomic invasions by exogenous FeLVs of the germ lines of *F. margarita* or *F. nigripes*.

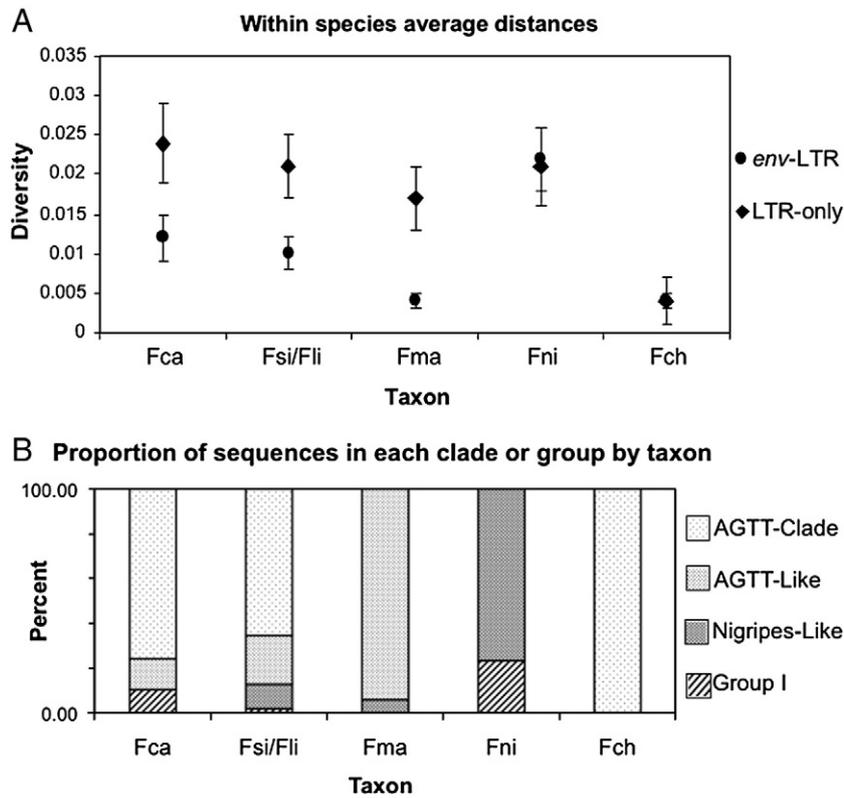
#### Comparisons of enFeLVs within and across species

While a “combined-LTR” phylogeny indicated clustering of sequences for *F. nigripes*, *F. chaus* and *F. margarita*, the dataset was too short in sequence length to robustly identify distinctive or well-supported clades. Thus this dataset was only used in non-phylogenetic analyses within and across species. Within species, *F. margarita* and *F. chaus* were found to have the lowest genetic diversity values, indicating a relatively high homogeneity of enFeLV sequences detected within these species (Fig. 3A). This is consistent with the clustering evident for these species in the *env*-LTR phylogeny (Fig. 2), in which all but one *F. margarita* sequences fell into a single clade, while *F. chaus* sequences were all similar in sequence to those of domestic and wildcat enFeLV-AGTT.

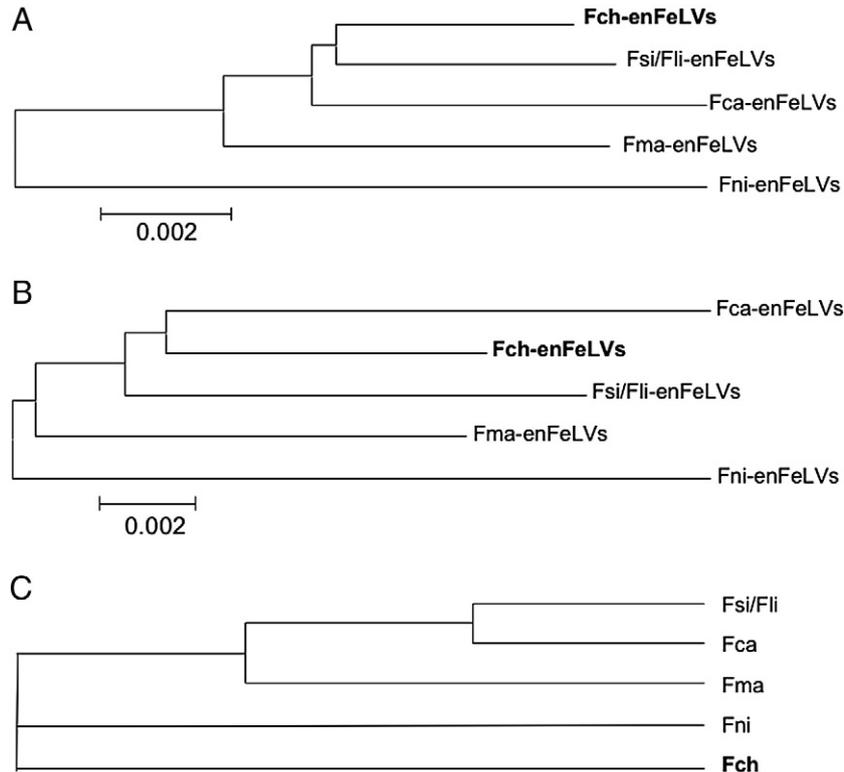
The diversity present for *env*-LTR within species (Fig. 3A) reflected the degree of clustering within the tree (Fig. 2), but would also be affected by the degree to which sequences were dispersed across the highly differentiated Groups I and II. Thus the distribution of *env*-LTR enFeLV sequences among Group I and among the subdivisions of Group II was determined for each species (Fig. 3B) and compared to the diversity of enFeLVs (Fig. 3A). Group I enFeLVs were not at all detected among the low-diversity jungle cat, *F. chaus*, which carried only AGTT-clade enFeLVs (Figs. 2, 3). Somewhat higher diversity was present in the sand cat, *F. margarita*, which lacked Group I sequences but had enFeLVs outside the *margarita*-clade, as well as enFeLV diversity within the *margarita*-clade (Figs. 2, 3). Higher enFeLV diversity values were estimated for the black-footed cat, *F. nigripes*, for which Group I and Group II sequences were common in the *env*-LTR dataset, and likewise for the domestic cat and wildcat (Figs. 2, 3).

To determine whether the patterns present among enFeLV sequences recapitulated the known species relationships, the species tree for the domestic cat lineage (Driscoll et al., 2007; Johnson et al., 2006) was compared to trees generated using a distance matrix of enFeLV sequences (Fig. 4). The distance matrix produced using enFeLV sequences from the *env*-LTR dataset and a second distance

**Fig. 2.** Phylogenetic relationships inferred for 130 sequences across species within the cat genus *Felis* using a 721 bp alignment spanning *env* to the 3' LTR region of enFeLV. The sequences shown were generated from cloned amplicons following PCR using the primer set *env*-LTR (Fig. 1; Table 1), and include two domestic cat sequences from GenBank (black font). The phylogeny was inferred with partitioned maximum likelihood. Most nodes did not receive strong support. The scale bar denotes 0.002 substitutions/site. The tree shown was midpoint-rooted.  $\text{LnLik} = -2795.068387$ ;  $\alpha_{\text{env}} = 0.130703$ ,  $\alpha_{39\text{bp}} = 0.801663$ ,  $\alpha_{\text{LTR}} = 0.198288$ . Sequence labels contain the species and sample number (see Table 1) followed by “3” (which designates the *env*-LTR primer set), followed by the clone number. Duplicate sequences within a species were included only once in the dataset and the number of duplicates is indicated after the name of the representative sequence (e.g. “X2”). Sequences with the 14 bp insertion are marked by asterisks. Sequence labels are colored by taxon: *F. catus* (violet), *F. lybica* (teal), *F. silvestris* (blue), *F. margarita* (green), *F. nigripes* (orange), *F. chaus* (mauve). The AGTT-clade was defined by the position of enFeLV-AGTT, a previously identified domestic cat enFeLV provirus that is genomically intact and of recent origin (Roca et al., 2004).



**Fig. 3.** Within-species diversity of enFeLVs. (A) Summary by taxon of the genetic diversity present among enFeLVs for both the *env*-LTR and LTR-only datasets. Diversity within the taxa was determined using MEGA. (B) The proportion of *env*-LTR sequences for each taxon that were in each of the clades or groups distinguished in the *env*-LTR phylogeny shown in Fig. 2; duplicate sequences were included in determining the totals.



**Fig. 4.** Relationships of enFeLV sequences present across taxa, as inferred using MEGA from a distance matrix generated between pairs of felid taxa using their enFeLV sequences. (A) Relationships across the enFeLV *env*-LTR sequences present across taxa. (B) Relationships across the enFeLV LTR-only sequences present across taxa. (C) Taxonomic relationships among species within the domestic cat lineage (Johnson et al., 2006); the species tree is a cladogram in which branch lengths are not proportional to genetic distances or divergence times. Both sets of enFeLV sequences recapitulate the relationships present among taxa within the domestic cat lineage for all species except the jungle cat *F. chaus*. Since *F. chaus* as a species is found at or close to a basal position within the domestic cat lineage, the discordant position inferred for that species by the distance matrices of enFeLV sequences suggests that a secondary and/or recent germ line invasion or proliferation of enFeLVs may have affected the jungle cat.

matrix produced using the enFeLV LTR-only dataset both generated trees with similar topology (Fig. 4), except that the positions of the domestic cat and its progenitor the wildcat are interchanged. These enFeLV trees matched the true species relationships known to be present within the domestic cat lineage with one exception (Fig. 4). The two most basal species within the domestic cat lineage are the jungle cat, *F. chaus*, and the black-footed cat, *F. nigripes* (though which of these two lineages diverged first has not been resolved) (Johnson et al., 2006). The *F. nigripes* enFeLV sequences did demonstrate the greatest degree of distance from those of other species in the lineage (Fig. 4), consistent with the basal position of the species as determined by previous mtDNA and nuclear genomic data (Johnson et al., 2006; Johnson and O'Brien, 1997; King et al., 2007; Pecon Slattery and O'Brien, 1998; Pecon Slattery et al., 2000; Pecon-Slattery et al., 2004). By contrast, the other basal species, *F. chaus*, did not appear to have highly divergent enFeLV sequences (Fig. 4), suggesting that the genome of this species may have been subject to a recent invasion or proliferation by FeLVs that did not affect *F. nigripes*, but that generated discordance between the actual species tree for the domestic cat lineage and the trees derived using enFeLV sequences (Fig. 4).

Finally, the fixation index, a measure of population differentiation, was estimated by comparing enFeLV sequences between taxa (Table 2). The highest  $F_{ST}$  values were estimated for comparisons that involved any two of the following species: *F. nigripes*, *F. margarita* and *F. chaus* (Table 2). This was consistent with the clustering and/or distinctiveness of the enFeLVs evident for these three species on the *env*-LTR phylogeny (Fig. 2). Additionally, since one hypothesis was that the process of cat domestication might have led to distinct patterns for enFeLVs in the domestic cat when compared to its wildcat ancestor (Roca et al., 2004), the results of the  $F_{ST}$  calculation between the domestic cat *F. catus* and the wildcat *F. silvestris*/*F. lybica* were especially meaningful. For both *env*-LTR and the LTR-only enFeLV sequences, the  $F_{ST}$  calculated between the domestic cat and the wildcat was lower than for any other comparison between taxa, and suggested that no significant differentiation was present in enFeLVs between domestic cat and wildcat (Table 2). Although in the *env*-LTR phylogeny, wildcat but not domestic cat sequences were present within the *nigripes*-like groups (Figs. 2, 3), the presence of domestic cat enFeLVs within these groups was suggested by BLAST queries (see above), so the apparent absence of domestic cat enFeLVs from this part of the phylogeny was likely due to limited sampling. The enFeLV sequences for the two taxa appear to be otherwise largely interspersed, and there was no distinctive large cluster of domestic cat enFeLVs that would distinguish them from the enFeLVs of the wildcat.

**Table 2**  
 $F_{ST}$  estimates using enFeLV sequences between species.

| Comparison           | <i>env</i> -LTR |                   | LTR-only       |                   |
|----------------------|-----------------|-------------------|----------------|-------------------|
|                      | $F_{ST}$ value  | P-value           | $F_{ST}$ value | P-value           |
| Fca – Fsi/Fli        | 0.044           | 0.0039            | 0 <sup>a</sup> | 0.3970            |
| Fca – Fma            | 0.226           | <0.0001           | 0.104          | 0.0137            |
| Fca – Fni            | 0.140           | <0.0001           | 0.086          | 0.0059            |
| <b>Fca – Fch</b>     | <b>0.158</b>    | <b>&lt;0.0001</b> | <b>0.173</b>   | <b>0.0020</b>     |
| Fsi/Fli – Fma        | 0.130           | <0.0001           | 0.059          | 0.0225            |
| <b>Fsi/Fli – Fni</b> | <b>0.152</b>    | <b>&lt;0.0001</b> | <b>0.126</b>   | <b>0.0010</b>     |
| Fsi/Fli – Fch        | 0.041           | 0.0469            | 0.087          | 0.0078            |
| <b>Fma – Fni</b>     | <b>0.170</b>    | <b>&lt;0.0001</b> | <b>0.206</b>   | <b>&lt;0.0001</b> |
| <b>Fma – Fch</b>     | <b>0.361</b>    | <b>&lt;0.0001</b> | <b>0.383</b>   | <b>&lt;0.0001</b> |
| <b>Fni – Fch</b>     | <b>0.192</b>    | <b>&lt;0.0001</b> | <b>0.349</b>   | <b>&lt;0.0001</b> |

Taxon designations follow those of Table 1.

Significance level was set to  $p < 0.0025$  after Bonferroni correction.

Species pairs with significantly high  $F_{ST}$  values for both sets of sequences are in bold.

<sup>a</sup> Due to imprecision in estimation,  $F_{ST}$  for Fca – Fsi/Fli was  $-0.002$ , and is set to null.

## Discussion

Endogenous FeLVs were amplified, cloned and sequenced for six species of the domestic cat lineage. A phylogeny based on a 722 bp proviral sequence of the *env*-LTR region of enFeLV (Table 1, Fig. 1) supported the subdivision of enFeLVs across the domestic cat lineage into two major clades, designated Groups I and II (Pontius et al., 2007; Roca et al., 2004), with additional subdivisions apparent within Group II enFeLVs (Fig. 2). Although novel deep lineages among enFeLVs in wild species of *Felis* were not detected, the current results nonetheless do suggest a complex history of interactions between FeLV and the germ lines of the species that comprise the domestic cat lineage. These findings are analogous to the different patterns seen across species of primates in studies involving endogenous retroviruses found in humans (Bannert and Kurth, 2006). In particular, the different enFeLV patterns detected across species of the domestic cat lineage likely resulted from invasion, proliferation or homogenization of enFeLVs that varied among host species of *Felis*.

The greatest differences detected among enFeLV sequences between taxa were found in comparisons that involved *F. nigripes*, *F. margarita* and *F. chaus* (Table 2). This was consistent with the clustering and/or distinctiveness of the enFeLVs evident for these three species in the *env*-LTR phylogeny (Fig. 2): all but one *F. margarita* enFeLV sequences clustered in the *margarita*-clade while all *F. chaus* sequences clustered within the separate AGTT-clade (Fig. 2). Although *F. nigripes* enFeLVs were diverse, *F. nigripes* sequences were absent from both the AGTT-clade and the “AGTT-like groups” category that contained the *margarita*-clade. This separation in the *env*-LTR phylogeny of enFeLV sequences for the three species would produce the high  $F_{ST}$  values that were obtained in comparisons between any two of these three species (Table 2). Thus both the phylogeny and diversity measurements appeared to reflect and be greatly affected by evolutionary events in the history of enFeLV that involved some species but not others. Although the effects of sampling bias cannot completely be ruled out, these results appear to reflect an abundance of very similar AGTT-clade enFeLV sequences in the jungle cat *F. chaus*; the absence of enFeLVs in the AGTT-clade and AGTT-like groups for the black-footed cat *F. nigripes*; and the presence of a seemingly monophyletic clade of enFeLVs within the sand cat, *F. margarita*. In the case of the sand cat, the abundance of *margarita*-clade enFeLVs suggests that the sand cat has experienced a species-specific proliferation of this type of enFeLV, or possibly homogenization of the enFeLV sequences through gene conversion (Hughes and Coffin, 2005).

For all but one of the species, the relationships determined using enFeLV sequences matched those of the species tree of the domestic cat lineage (Fig. 4). The only exception was the jungle cat *F. chaus*, in which only Group II enFeLVs within the derived AGTT-clade were detected. Nonetheless, it seems likely that the absence of diverse enFeLVs in *F. chaus* in the *env*-LTR dataset may not reflect the true diversity of enFeLVs present in the germ line of that species, for several reasons: (1) fewer sequences of enFeLV were successfully produced for *F. chaus* than for any other species, therefore adversely affecting our inference power on this species; (2) it appears that in *F. chaus* there was a substantial recent proliferation of AGTT-clade enFeLVs, and that the large number of copies of AGTT-clade enFeLVs may have made it difficult to detect other types of enFeLVs that may have been present in the species; (3) in other species in which a similar proliferation of a particular type of sequence is evident, such as *F. margarita*, it was difficult to detect the presence of enFeLVs from other clades; (4) finally, it has not been possible to determine whether *F. chaus* or *F. nigripes* is more basal among species in the domestic cat lineage based on genomic analysis (Johnson et al., 2006), suggesting that divergence of one of their lineages was quickly followed by divergence of the second lineage, allowing for only a short evolutionary interval during which FeLV could have invaded the germ line of the last common ancestor of *F. nigripes* and the other

species but not *F. chaus* (assuming that *F. chaus* was actually the most basal species in the domestic cat lineage). It cannot be completely ruled out that if *F. chaus* were basal to the species tree of *Felis*, the original invasion by enFeLV of the domestic cat lineage germ line may have occurred after *F. chaus* diverged, with a secondary and more recent invasion of its intact germ line. Although the current study cannot rule out this last possibility, the most parsimonious explanation for the lack of Group I or *nigripes*-like sequences among *F. chaus* is that a recent invasion or proliferation of AGTT-clade enFeLVs has eclipsed the older proviral diversity that may be present within the jungle cat. Finally, while it is possible that some species were more greatly affected than others by demographic effects (drift, bottlenecks, and founder effects), these would have tended to alter diversity within a locus, and caused insertionally polymorphic loci to be removed or fixed in the species; they would not have led to changes in the types of loci present in the species (Macfarlane and Simmonds, 2004).

Since the *F. chaus* env-LTR sequences are similar to domestic cat enFeLV-AGTT (Fig. 2), it is possible that if a recent invasion of the *F. chaus* germ line occurred, it may have been due to hybridization with or infection by the wildcat. The enFeLV sequences within the AGTT-clade show greater diversity in the domestic cat and wildcat than in the jungle cat, hinting at an earlier origin for enFeLV-AGTT clade proviruses in the wildcat than in the jungle cat. The domestic cat enFeLV-AGTT (which defines and is basal to the clade that bears its name) appears to be of recent origin and genomically intact (although replication competence was not established) (Roca et al., 2004). Although the effects of sampling bias cannot completely be ruled out, the absence of AGTT-clade sequences in *F. nigripes* and (possibly) *F. margarita* (Fig. 2) is further evidence that the AGTT-clade may represent a proliferation of enFeLVs in the wildcat germline that may have given rise to a secondary invasion of the genome of *F. chaus*. This would also explain why *F. chaus* was the only species for which the enFeLV sequence relationships did not recapitulate the species relationships within the domestic cat lineage (Fig. 4).

Among the species examined, the black-footed cat, *F. nigripes*, had a very high level of enFeLV diversity within both analyzed datasets (env-LTR and LTR-only) (Fig. 3A). This high diversity was likely due both to the high proportion of Group I sequences relative to other species (Fig. 3B) and to the high diversity present among sequences of *F. nigripes* within each of the major clades (Fig. 2). The distinctiveness of enFeLVs in *F. nigripes* likely reflected the ancient origin of its enFeLVs and its status as one of the first lineages within *Felis* to diverge. Additionally, one of the most striking aspects of the distribution of enFeLVs across species was the complete absence of *F. nigripes* enFeLVs within the AGTT-clade and AGTT-like groups that comprise most of Group II (Fig. 2). There are a sufficient number of sequences of enFeLVs from *F. nigripes* to conclude that this observed distribution is representative of the species and not due to inadequate sequencing (Fig. 2). This pattern would in turn have produced for *F. nigripes* the relatively higher proportions of enFeLV sequences present in Group I and in the Group II “*nigripes*-like-groups” to which the species therefore gave its name. The distribution of *F. nigripes* enFeLVs within Group II also indicates that Group II enFeLVs invaded or proliferated in the germ line of the domestic cat lineage in two stages, one preceding and one subsequent to the divergence of the lineage of *F. nigripes* from the ancestor of all modern species of *Felis*.

The current analyses do not support the hypothesis that highly distinctive clades of enFeLV may be present in wild species that are not present in the domestic cat (Fig. 2). The domestic cat (*F. catus*) and its immediate progenitor, the wildcat (*F. silvestris*/*F. lybica*), demonstrated a tremendous diversity of enFeLVs, with sequences interspersed throughout the phylogeny (Fig. 2). Domestication has greatly increased the density at which domestic cats live, relative to the density of wild cats, and exogenous FeLV infections are an acute problem for cats that are crowded together (Lee et al., 2002). Since exFeLVs interact with enFeLVs, one hypothesis was that strong selection for or against certain

enFeLVs might have affected only domestic cats. The current results do not rule out a role for selection on individual copies of enFeLV in the domestic cat. Nonetheless, there are no distinctive enFeLV groups that are found only in domestic cats to the exclusion of the wildcat. Our current evidence suggests that any changes in the pattern of enFeLVs in the domestic cat relative to its wildcat progenitors were not pervasive, and that a distinctive or widespread invasion or proliferation of enFeLVs has not occurred since the cat was domesticated.

## Materials and methods

### Species and breeds

Samples were collected from male individuals from a diverse set of species, subspecies and geographic locations (Table 1). All samples were from the collection at the Laboratory of Genomic Diversity (LGD), National Cancer Institute (NCI), Frederick, Maryland, USA. Samples included white blood cells, skin, spleen or cell lines. DNA extractions were attempted from all seven species of the domestic cat lineage (genus *Felis*) known to carry enFeLVs (Table 1). Individuals on which previous enFeLV or other research has been conducted were preferentially selected, to allow potential comparison of results across studies, including Fca-146 (Pontius et al., 2007; Roca et al., 2005, 2004), Fca-9 (Roca et al., 2005), Fca-127 (Roca et al., 2005, 2004), Fsi-1 (Pecon Slattery and O'Brien, 1998), Fsi-5 (Driscoll et al., 2007), Fsi-13 (Driscoll et al., 2007), Fma-8 (Johnson and O'Brien, 1997; Pecon Slattery and O'Brien, 1998), Fma-10 (Johnson and O'Brien, 1997), Fni-6 (Johnson and O'Brien, 1997), and Fni-14 (Johnson and O'Brien, 1997; Pecon Slattery and O'Brien, 1998).

Asian wildcats (sometimes classified separately as *Felis ornata*), along with the European wildcat and the African wildcat are often considered to form a single interbreeding species (Driscoll et al., 2007), and assignment of wildcats to one of the two (or three) species is often inconsistent even for wildcats from nearby geographic locations (Table 1). Thus for this study, all wildcat samples, whether designated *F. lybica* or *F. silvestris*, were combined for all analyses. The domestic cat is sometimes considered to be the same species as the wildcat *Felis lybica* (Driscoll et al., 2007); nonetheless, since one of the goals of the study was to examine differences between domestic and wild cats in enFeLVs, the domestic cats were analyzed separately from their wildcat progenitors.

### DNA extraction and screening

DNA was extracted from all samples using the DNeasy kit (Qiagen). Unlike most of the proviral enFeLV genome, the U3 region of the LTR is not homologous between enFeLVs and exFeLVs (Berry et al., 1988). Thus all DNA samples were tested using PCR to rule out individuals carrying exogenous FeLV proviruses, using published primers for the U3 region of the LTRs specific only for exogenous FeLV and not endogenous FeLV (I2-xU3-F1 and I2-xU3-R1) (Roca et al., 2005), with the DNA quality of the same samples verified using primers PFL-F4 and PFL-R4 that are designed to amplify only endogenous and not exogenous FeLVs (Roca et al., 2005).

### Amplification of enFeLV

The enFeLV provirus includes two LTR regions (which flank the protein coding genes) that are identical at the time of integration (Johnson and Coffin, 1999). Thus, both the 5' and the 3' LTRs would be simultaneously amplified by polymerase chain reaction (PCR) using a primer pair internal to the LTR. These LTR primers were designed using conserved regions of previously sequenced enFeLV LTRs in domestic cats (GenBank accession numbers: AY364318-9, AY364320-2, M18247, M19392, AF052723) (Chen et al., 1998; Donahue et al., 1988; Roca et al., 2005, 2004). The other primer pair, designated *env*-

LTR, had a forward primer based on enFeLV *env* sequence and a reverse primer based on the enFeLV LTR region, and thus would only amplify 3' LTRs in proviruses with a coding region adjacent. Primer sequences and locations are shown in Fig. 1.

Each PCR setup included at least one domestic cat with good quality DNA (as a positive control for enFeLV), and two negative controls, of the extraction and amplification, to ensure lack of contamination. All reagents used double-distilled water exposed to UV radiation to minimize the risk of contamination of PCR mixtures. Two separate PCRs were performed for each individual for each primer pair, in order to increase the possibility of detecting the genetic diversity present in the individuals. All PCRs were performed in a total volume of 25  $\mu$ l including 10 $\times$  PCR buffer, 0.2 mM of dNTPs, 1.5 mM of MgCl<sub>2</sub>, 0.4  $\mu$ M of each primer and 0.5 U/reaction of AmpliTaq Gold [Applied Biosystems, Inc (ABI)] using a touchdown PCR method consisting of an initial denaturation at 95 °C for 10 min followed by a total of 45 cycles of 15 s denaturation at 94 °C, 45 s annealing for two cycles each at 60 °C, 58 °C, 56 °C, 54 °C, 52 °C, and 35 cycles at 50 °C or 48 °C, and 45 s elongation at 72 °C, with a final elongation step of 10 min at 72 °C. The PCRs were performed using high fidelity AmpliTaq Gold to minimize polymerase errors. PCR products were eluted using electrophoresis, purified using Microcon-50 filter devices (Millipore), with preliminary direct sequencing of product in order to verify the identity of the region amplified as the target region, followed by cloning.

#### Cloning and sequencing of enFeLV amplifications

To examine the diversity of sequences within each sample, purified PCR products were cloned into a TOPO-TA vector and grown in One Shot TOP10 chemically competent *E. coli* cells (Invitrogen). At least 8–12 isolated colonies from each transformation were grown, with DNA isolated using a QIAprep kit (Qiagen). Sequences were generated using the BigDye Terminator Cycle Sequencing kit (ABI) and resolved on an ABI PRISM 3700 DNA Analyzer. For each positive clone, primers kz64 and kz77 (Weber et al., 2004) were used for sequencing the insert. Although the possibility of sequencing errors cannot be completely excluded, this possibility was minimized by the generation of overlapping sequences for each clone using both sense and anti-sense primers.

#### Sequence verification and alignment

The sequences obtained were initially verified as enFeLVs using the NCBI BLAST algorithm (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>). Chromatograms were individually examined in order to confirm the quality of sequences, using Sequencher 4.8 (Gene Codes Corp.). The forward and reverse sequences of each clone were assembled into a contig in Sequencher. Each individual contig was visually inspected and verified, any ambiguities were visually resolved, and a final sequence was generated for further analyses. Sequences with poor quality chromatograms were excluded from the study. Sequences were compared to one exFeLV and two enFeLVs that were used as reference sequences (exFeLV: M25425; and two enFeLVs: AY364318.1, AY364319.1) (Kumar et al., 1989; Roca et al., 2004). All sequences were deposited in GenBank (accession numbers: GU300813–GU301075).

Diagnostic sites for clades consisting of enFeLV sequences only from *F. margarita* or only from *F. nigripes* were queried using NCBI Megablast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) against trace archives of the domestic cat genome (<http://www.ncbi.nlm.nih.gov/Traces/trace.cgi>).

#### Phylogenetic analyses

Within-individual clones that produced identical sequences may have been due to repeated amplification of the same locus within an

individual sample. Thus where two within-individual clones had identical sequences, duplicate sequences were removed, with only one of the sequences chosen for the phylogeny as representative of other duplicate sequences in that individual. For the *env*-LTR dataset: Fca127-3-4 was representative of Fca-127-3-9; Fca-146-3-1 of Fca-146-3-6; Fli-3-3-1 of Fli-3-3-2, Fli-3-3-3 and Fli-3-3-7; Fsi-7-3-2 of Fsi-7-3-12, Fsi-7-3-9 and Fsi-7-3-5; Fsi-9-3-2 of Fsi-9-3-3 and Fsi-9-3-4; Fsi-24-3-8 of Fsi-24-3-10; Fma-8-3-3 of Fma-8-3-6; Fma-8-3-9 of Fma-8-3-10; Fma-10-3-1 of Fma-10-3-3; Fma-10-3-2 of Fma-10-3-6; Fni-15-3-2 of Fni-15-3-10; Fch-1-3-7 of Fch-1-3-9 and Fch-1-3-10. For the LTR-only dataset: Fch-1-2-3 was representative of Fch-1-2-2, Fch-1-2-5, Fch-1-2-6, Fch-1-2-7, Fch-1-2-9, Fch-1-2-10, Fch-1-2-11 and Fch-1-2-12; Fch-11-2-6 of Fch-11-2-2, Fch-11-2-3, Fch-11-2-5, Fch-11-2-7, Fch-11-2-9, Fch-11-2-10 and Fch-11-2-11; Fsi-1-2-8 of Fsi-1-2-10; Fsi-5-2-5 of Fsi-5-2-6; Fsi-5-2-3 of Fsi-5-2-10 and Fsi-5-2-12; Fsi-13-2-8 of Fsi-13-2-5 and Fsi-13-2-9; Fsi-24-2-5 of Fsi-24-2-8. Moreover, for the combined *env*-LTR and LTR-only analyses, identical sequences with 100% identity between the two datasets were checked for within each individual, and a total of 12 additional duplicates were removed, with only one of the sequences chosen for the phylogeny as representative of the duplicates: Fca-9-3-6 was representative of Fca-9-2-4; Fca-9-3-11 of Fca-9-2-5; Fch-11-3-10 of Fch-11-2-6X8 (which represents 8 previously identified sequences); Fch-1-3-1 of Fch-1-2-3X9; Fma-8-3-4 of Fma-8-2-11; Fma-8-3-12 of Fma-8-2-6; Fni-6-3-1 of Fni-6-2-10; Fsi-13-3-9 of Fsi-13-2-6; Fsi-13-3-3 of Fsi-13-2-8X3; Fsi-1-3-4 of Fsi-1-2-8X2; Fsi-24-3-9 of Fsi-24-2-5X2; Fsi-24-3-7 of Fsi-24-2-3. Finally, two sequences were excluded from the *env*-LTR data set (Fca-146-3-4; Fsi-9-3-10) due to their short lengths (255 and 413 bp in the final alignment) relative to the other sequences (722 bp).

We investigated the phylogenetic history of FeLV sequences using maximum likelihood (ML) inference and the GTR substitution model (Lanave et al., 1984) with rate heterogeneity modeled by the  $\Gamma$  distribution and four rate categories (Yang, 1994). The alignment was partitioned in three loci: *env*, the 39-nt spacer, and the 3' LTR. Ten independent tree searches were launched using independently estimated stepwise-addition parsimony starting trees. The ninth search yielded the best likelihood value and was used to further optimize the topology, branch lengths, and model parameters. Node support was evaluated with 1500 bootstrap pseudo-replicates (Felsenstein, 1985) and the final bipartitions were drawn on the best likelihood tree, thus expressing the proportion of bootstrap trees containing the clades of the best ML tree. All ML phylogenetic analyses were carried out in RAXML 7.2.5 (Stamatakis, 2006) making use of its Pthreads version (Stamatakis and Ott, 2008) on multi-core computer architecture.

#### Recombination detection

Among endogenous FeLVs, recombination may occur between 5' and 3' LTRs, giving rise to solo LTRs (Casey et al., 1981); while recombination including gene conversion is known to affect endogenous retroviruses in general (Hughes and Coffin, 2005). We therefore tested for the presence of breakpoints and recombination in the enFeLV datasets. The recombination scans were performed in RDP 3b41 (Martin et al., 2005) by dividing the *env*-LTR dataset into three segments (*env*, 39-nt spacer, and 3' LTR segments) and using methods including RDP (Martin and Rybicki, 2000; Martin et al., 2005), GeneConv (Padidam et al., 1999), Topal (McGuire and Wright, 2000), MaxChi (Maynard Smith, 1992), and concatenated using the abovementioned algorithms and the modified bootscanning method (Martin et al., 2005). Bootscanning was performed with phylogenetic trees estimated with the Neighbor Joining method (Saitou and Nei, 1987) and the F84 substitution model (Felsenstein, 1984), 500 nonparametric bootstrap pseudo-replicates, an alignment window size of 200

nucleotides and step size of 20–50 nucleotides; the transition–transversion ratio ( $R = 4.3$ ) was estimated in MEGA 4.1 (Tamura et al., 2007).

The existence of sequence features such as indels in phylogenetically distant sequences may hint at recombination. We detected a 19-nt deletion present in all but one Group I enFeLV sequences and present in Group II only for three *F. nigripes* enFeLV sequences from the *nigripes*-like group. In order to understand whether the deletion originated from a single event or multiple ones, we mapped the indel events on the ML tree using parsimony and likelihood reconstruction methods as implemented in Mesquite 2.72 (<http://mesquiteproject.org>). The parsimony method attempts to reconstruct the ancestral state of the indel at each node minimizing its state changes along the phylogeny, while the likelihood methods does the same by maximizing the probability of evolution of the observed indel state (i.e. insertion vs. deletion). The Mk1 (Markov  $k$ -state 1-parameter) evolutionary model was employed for the likelihood reconstruction.

#### Population genetic analyses

Genetic diversity of enFeLV was calculated for each host taxon in MEGA (Tamura et al., 2007), using 128 and 135 sequences for *env*-LTR and LTR-only data, respectively. The mean diversity was calculated for each taxon using the Kimura 2-parameter substitution model (Kimura, 1980) such that:

$$\pi_i = \frac{q}{q-1} \sum_{i=1}^q \sum_{j=1}^q x_i x_j d_{ij}$$

where  $x_i$  is the frequency of  $i$ -th sequence in the sample from taxon  $i$ ;  $q$  is the number of different sequences from each taxon; and  $d_{ij}$  is the pairwise distance between sequences  $i$  and  $j$  (Tamura et al., 2007).

In order to quantify the apportionment of viral genetic variance across hierarchically organized sets of sequences, we grouped enFeLV sequences as above by host taxon and performed an analysis of molecular variance. This population genetic analysis would be complicated by the following factors: (i) the use of multiple enFeLV loci rather than a single locus; (ii) the potential for the same sequence to exist at more than one locus; (iii) the possibility that the same locus could be sequenced multiple times within an individual; and (iv) stochastic effects leading to the sequencing of different loci across individuals or species. Nonetheless, since these features would impact all species and individuals, the analyses would be meaningful in terms of relatively higher or lower values generated for different pairwise comparisons.  $F_{ST}$  was calculated for all possible pairs of taxa based on the enFeLV sequences for each taxon. These analyses were carried out in Arlequin 3.11 (Excoffier et al., 2005). Since a total of twenty hypotheses were tested (ten for each of the two sequence datasets), a Bonferroni correction was applied for multiple comparisons, and the level of statistical significance was set to 0.0025.

#### Acknowledgments

We thank S. J. O'Brien and the LGD/NCI for initial access to samples. All samples had been collected in compliance with federal regulations and permits. This work was supported by the Morris Animal Foundation grant D06FE-029. SOK was supported by the U.S. Defense Advanced Research Projects Agency (DARPA).

#### References

Bannert, N., Kurth, R., 2006. The evolutionary dynamics of human endogenous retroviral families. *Annu. Rev. Genomics Hum. Genet.* 7, 149–173.  
 Benveniste, R.E., Sherr, C.J., Todaro, G.J., 1975. Evolution of type C viral genes: origin of feline leukemia virus. *Science* 190 (4217), 886–888.  
 Berry, B.T., Ghosh, A.K., Kumar, D.V., Spodick, D.A., Roy-Burman, P., 1988. Structure and function of endogenous feline leukemia virus long terminal repeats and adjoining regions. *J. Virol.* 62 (10), 3631–3641.

Casey, J.W., Roach, A., Mullins, J.L., Burck, K.B., Nicolson, M.O., Gardner, M.B., Davidson, N., 1981. The U3 portion of feline leukemia virus DNA identifies horizontally acquired proviruses in leukemic cats. *Proc. Natl. Acad. Sci. U. S. A.* 78 (12), 7778–7782.  
 Chen, H., Bechtel, M.K., Shi, Y., Phipps, A., Mathes, L.E., Hayes, K.A., Roy-Burman, P., 1998. Pathogenicity induced by feline leukemia virus, Rickard strain, subgroup A plasmid DNA (pFRA). *J. Virol.* 72 (9), 7048–7056.  
 Coffin, J.M., 2004. Evolution of retroviruses: fossils in our DNA. *Proc. Am. Philos. Soc.* 148 (3), 264–280.  
 Culver, M., Menotti-Raymond, M.A., O'Brien, S.J., 2001. Patterns of size homoplasy at 10 microsatellite loci in pumas (*Puma concolor*). *Mol. Biol. Evol.* 18 (6), 1151–1156.  
 Donahue, P.R., Hoover, E.A., Beltz, G.A., Riedel, N., Hirsch, V.M., Overbaugh, J., Mullins, J.L., 1988. Strong sequence conservation among horizontally transmissible, minimally pathogenic feline leukemia viruses. *J. Virol.* 62 (3), 722–731.  
 Driscoll, C.A., Menotti-Raymond, M., Roca, A.L., Hupe, K., Johnson, W.E., Geffen, E., Harley, E.H., Delibes, M., Pontier, D., Kitchener, A.C., Yamaguchi, N., O'Brien, S.J., Macdonald, D.W., 2007. The Near Eastern origin of cat domestication. *Science* 317 (5837), 519–523.  
 Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1, 47–50.  
 Felsenstein, J., 1984. Distance methods for inferring phylogenies: a justification. *Evolution* 38, 16–24.  
 Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.  
 Flint, S.J., Enquist, L.W., Racaniello, V.R., Shalka, A.M., 2004. Principles of Virology – Molecular Biology, Pathogenesis, and Control of Animal Viruses, Second ed. ASM Press, USA.  
 Hughes, J.F., Coffin, J.M., 2005. Human endogenous retroviral elements as indicators of ectopic recombination events in the primate genome. *Genetics* 171 (3), 1183–1194.  
 Johnson, W.E., Coffin, J.M., 1999. Constructing primate phylogenies from ancient retrovirus sequences. *Proc. Natl. Acad. Sci. U. S. A.* 96 (18), 10254–10260.  
 Johnson, W.E., O'Brien, S.J., 1997. Phylogenetic reconstruction of the Felidae using 16S rRNA and NADH-5 mitochondrial genes. *J. Mol. Evol.* 44 (Suppl 1), S98–S116.  
 Johnson, W.E., Eizirik, E., Pecon-Slatery, J., Murphy, W.J., Antunes, A., Teeling, E., O'Brien, S.J., 2006. The late Miocene radiation of modern Felidae: a genetic assessment. *Science* 311 (5757), 73–77.  
 Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120.  
 King, V., Goodfellow, P.N., Pearks Wilkerson, A.J., Johnson, W.E., O'Brien, S.J., Pecon-Slatery, J., 2007. Evolution of the male-determining gene SRY within the cat family Felidae. *Genetics* 175 (4), 1855–1867.  
 Koshy, R., Gallo, R.C., Wong-Staal, F., 1980. Characterization of the endogenous feline leukemia virus-related DNA sequences in cats and attempts to identify exogenous viral sequences in tissues of virus-negative leukemic animals. *Virology* 103 (2), 434–445.  
 Kumar, D.V., Berry, B.T., Roy-Burman, P., 1989. Nucleotide sequence and distinctive characteristics of the *env* gene of endogenous feline leukemia provirus. *J. Virol.* 63 (5), 2379–2384.  
 Lanave, C., Preparata, G., Saccone, C., Serio, G., 1984. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* 20, 86–93.  
 Lee, I.T., Levy, J.K., Gorman, S.P., Crawford, P.C., Slater, M.R., 2002. Prevalence of feline leukemia virus infection and serum antibodies against feline immunodeficiency virus in unowned free-roaming cats. *J. Am. Vet. Med. Assoc.* 220, 620–622.  
 Macfarlane, C., Simmonds, P., 2004. Allelic variation of HERV-K(HML-2) endogenous retroviral elements in human populations. *J. Mol. Evol.* 59 (5), 642–656.  
 Martin, D., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16, 562–563.  
 Martin, D.P., Williamson, C., Posada, D., 2005. RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 21, 260–262.  
 Maynard Smith, J., 1992. Analyzing the mosaic structure of genes. *J. Mol. Evol.* 34, 126–129.  
 McDougall, A.S., Terry, A., Tzavaras, T., Cheney, C., Rojko, J., Neil, J.C., 1994. Defective endogenous proviruses are expressed in feline lymphoid cells: evidence for a role in natural resistance to subgroup B feline leukemia viruses. *J. Virol.* 68 (4), 2151–2160.  
 McGuire, G., Wright, F., 2000. TOPAL 2.0: improved detection of mosaic sequences within multiple alignments. *Bioinformatics* 16, 130–134.  
 Mullins, J.L., Hoover, E.A., 1990. Molecular aspects of feline leukemia virus pathogenesis. In: Gallo, R.C., Wong-Staal, F. (Eds.), *Retrovirus Biology and Human Disease*. Marcel Dekker Inc., New York.  
 Niman, H.L., Akhavi, M., Gardner, M.B., Stephenson, J.R., Roy-Burman, P., 1980. Differential expression of two distinct endogenous retrovirus genomes in developing tissues of the domestic cat. *J. Natl. Cancer Inst.* 64 (3), 587–594.  
 Overbaugh, J., Riedel, N., Hoover, E.A., Mullins, J.L., 1988. Transduction of endogenous envelope genes by feline leukaemia virus in vitro. *Nature* 332 (6166), 731–734.  
 Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* 265, 218–225.  
 Pecon Slatery, J., O'Brien, S.J., 1998. Patterns of Y and X chromosome DNA sequence divergence during the Felidae radiation. *Genetics* 148 (3), 1245–1255.  
 Pecon Slatery, J., Sanner-Wachter, L., O'Brien, S.J., 2000. Novel gene conversion between X-Y homologues located in the nonrecombining region of the Y chromosome in Felidae (Mammalia). *Proc. Natl. Acad. Sci. U. S. A.* 97 (10), 5307–5312.  
 Pecon-Slatery, J., Pearks Wilkerson, A.J., Murphy, W.J., O'Brien, S.J., 2004. Phylogenetic assessment of introns and SINES within the Y chromosome using the cat family Felidae as a species tree. *Mol. Biol. Evol.* 21 (12), 2299–2309.

- Pontius, J.U., Mullikin, J.C., Smith, D.R., Lindblad-Toh, K., Gnerre, S., Clamp, M., Chang, J., Stephens, R., Neelam, B., Volfovsky, N., Schaffer, A.A., Agarwala, R., Narfstrom, K., Murphy, W.J., Giger, U., Roca, A.L., Antunes, A., Menotti-Raymond, M., Yuhki, N., Pecon-Slattery, J., Johnson, W.E., Bourque, G., Tesler, G., O'Brien, S.J., 2007. Initial sequence and comparative analysis of the cat genome. *Genome Res.* 17 (11), 1675–1689.
- Rezanka, L.J., Rojko, J.L., Neil, J.C., 1992. Feline leukemia virus: pathogenesis of neoplastic disease. *Cancer Invest.* 10 (5), 371–389.
- Roca, A.L., Pecon-Slattery, J., O'Brien, S.J., 2004. Genomically intact endogenous feline leukemia viruses of recent origin. *J. Virol.* 78 (8), 4370–4375.
- Roca, A.L., Nash, W.G., Menninger, J.C., Murphy, W.J., O'Brien, S.J., 2005. Insertional polymorphisms of endogenous feline leukemia viruses. *J. Virol.* 79 (7), 3979–3986.
- Roy-Burman, P., 1995. Endogenous env elements: partners in generation of pathogenic feline leukemia viruses. *Virus Genes* 11 (2–3), 147–161.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.
- Stamatakis, A., Ott, M., 2008. Efficient computation of the phylogenetic likelihood function on multi-gene alignments and multi-core architectures. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3977–3984.
- Stewart, M.A., Warnock, M., Wheeler, A., Wilkie, N., Mullins, J.I., Onions, D.E., Neil, J.C., 1986. Nucleotide sequences of a feline leukemia virus subgroup A envelope gene and long terminal repeat and evidence for the recombinational origin of subgroup B viruses. *J. Virol.* 58 (3), 825–834.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24 (8), 1596–1599.
- Tzavaras, T., Stewart, M., McDougall, A., Fulton, R., Testa, N., Onions, D.E., Neil, J.C., 1990. Molecular cloning and characterization of a defective recombinant feline leukaemia virus associated with myeloid leukaemia. *J. Gen. Virol.* 71 (Pt 2), 343–354.
- Weber, D.S., Stewart, B.S., Schienman, J., Lehman, N., 2004. Major histocompatibility complex variation at three class II loci in the northern elephant seal. *Mol. Ecol.* 13 (3), 711–718.
- Weiss, R.A., 2006. The discovery of endogenous retroviruses. *Retrovirology* 3, 67.
- Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* 39, 306–314.