



**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**  
**FACULDADE DE VETERINÁRIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS VETERINÁRIAS**  
**TESE DE DOUTORADO**

**HELENA MATA**

**EVOLUÇÃO E DIVERSIDADE DE RETROVÍRUS**  
**ENDÓGENOS EM FELÍDEOS NEOTROPICAIS**

**PORTO ALEGRE**

**2012**

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**  
**FACULDADE DE VETERINÁRIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS VETERINÁRIAS**  
**TESE DE DOUTORADO**

**EVOLUÇÃO E DIVERSIDADE DE RETROVÍRUS**  
**ENDÓGENOS EM FELÍDEOS NEOTROPICAIS**

Autor: HELENA MATA

Tese apresentada como requisito para obtenção do  
título de Doutor em Ciências Veterinárias  
na área de Microbiologia, especialidade Virologia.

Orientadora: Prof. Dra. Ana Paula Ravazzolo

PORTO ALEGRE

2012

CIP - Catalogação na Publicação

Mata, Helena  
EVOLUÇÃO E DIVERSIDADE DE RETROVÍRUS ENDÓGENOS EM  
FELÍDEOS NEOTROPICAIS / Helena Mata. -- 2012.  
110 f.

Orientadora: Ana Paula Ravazzolo.

Tese (Doutorado) -- Universidade Federal do Rio  
Grande do Sul, Faculdade de Veterinária, Programa de  
Pós-Graduação em Ciências Veterinárias, Porto Alegre,  
BR-RS, 2012.

1. Retrovírus endógenos. 2. felídeos neotropicais.  
3. Gammaretrovirus. I. Ravazzolo, Ana Paula, orient.  
II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os  
dados fornecidos pelo(a) autor(a).

Helena Mata

Evolução e diversidade de retrovírus endógenos em felídeos neotropicais.

Aprovada em 28 FEV 2012.

Este estudo foi realizado no Laboratório de Imunologia e Biologia Molecular do Departamento de Patologia Clínica Veterinária da Faculdade de Veterinária, Universidade Federal do Rio Grande do Sul, com apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico.

**À MINHA FAMÍLIA E AMIGOS**

## **AGRADECIMENTOS**

Agradeço a todas as pessoas que passaram pela minha vida e contribuíram para a minha formação profissional; sem citar nomes, para não correr o risco de esquecer alguém.

Agradeço, em especial, aos retrovírus endógenos que me ensinaram, através de suas histórias evolutivas, que só os fortes e persistentes atingem o sucesso.

A todos, muito obrigada!

## RESUMO

Retrovírus endógenos (ERVs) são vírus altamente difundidos no genoma de vertebrados. ERVs surgem quando retrovírus exógenos infectam células germinativas e se disseminam no genoma de seus hospedeiros, transmitindo seu material genético através das gerações por meio de herança mendeliana. ERVs são fundamentais na evolução dos genomas, sendo eles responsáveis por uma parte da diversidade genética de seus hospedeiros. O conhecimento sobre ERVs na família Felidae (Mammalia, Carnivora) estava praticamente restrito ao gato doméstico, e não se conhecia diversidade e padrões de evolução desses retroelementos em outras espécies. Este estudo teve como objetivo investigar diversidade, distribuição e padrões evolutivos de ERVs em espécies de gatos silvestres. Utilizando ferramentas de biologia molecular e bioinformática, foram identificadas e caracterizadas 85 sequências similares a retrovírus endógenos nos representantes das oito espécies brasileiras: *Leopardus pardalis*, *L. wiedii*, *L. colocolo*, *L. geoffroyi*, *L. tigrinus*, *Puma concolor*, *P. yagouaroundi* e *Panthera onca*. Encontrou-se uma predominância de ERVs similares a *Gammaretrovirus*, um padrão característico em muitas espécies de mamíferos. As análises filogenéticas evidenciaram três grupos principais de *Gammaretrovirus*, cada um evoluindo de maneira peculiar. Em uma visão geral, os ERVs provenientes de diferentes hospedeiros apresentaram-se distribuídos de forma heterogênea nas filogenias, dificultando a constatação de um padrão coevolutivo. No entanto, análises mais detalhadas de algumas sequências demonstraram peculiaridades, como no caso de um grupo de sequências similares a de um ERV oriundo do morcego *Myotis lucifugus*. Através de análises filogenéticas em comparação com dados obtidos na literatura, sugere-se que a infecção desse retrovírus ocorreu em uma espécie ancestral de felídeo, na segunda metade do Mioceno. Os resultados obtidos permitiram demonstrar que os felídeos neotropicais apresentam ERVs que seguem padrões semelhantes aos descritos a respeito de outros mamíferos, sugerindo também alguns casos de infecções de retrovírus muito similares entre diferentes ordens de mamíferos.

Palavras chaves: Retrovírus endógenos, felídeos neotropicais, mamíferos, *Gammaretrovirus*, evolução molecular.



## ABSTRACT

Endogenous retroviruses (ERVs) are widespread viruses in vertebrate genome. ERVs arise when exogenous retrovirus infects germinal cells and spread in the genome of their hosts, transmitting its genetic material throughout the generations by means of Mendelian inheritance. ERVs are fundamental for the evolution of genomes, being responsible for some part of the genetic diversity of their hosts. The knowledge on ERVs in felids (Mammalia, Carnivora, Felidae) was basically restricted to domestic cats, and the diversity and patterns of evolution of these retroviral elements in other species were not known. This study aimed to investigate diversity, distribution and evolutionary patterns of ERVs in wildcat species. Hence, by utilizing molecular biology and bioinformatics tools, 85 endogenous retrovirus-like sequences were identified and characterized in eight representative Brazilian species: *Leopardus pardalis*, *L. wiedii*, *L. colocolo*, *L. geoffroyi*, *L. tigrinus*, *Puma concolor*, *P. yagouaroundi* and *Panthera onca*. The analyses of these novel felid ERVs showed the predominance of *Gammaretrovirus*-like sequences, which is a characteristic pattern present in many mammal species. Phylogenetic analyses have evidenced three major groups of *Gammaretrovirus*, each one evolving in a peculiar manner. ERVs from different hosts were distributed in a mixed way in the phylogenies, differently of a coevolutionary pattern. However, more detailed analyses of some sequences demonstrated peculiarities, as in the case of a group of sequences similar to an ERV from the bat *Myotis lucifugus*. Notably, through phylogenetic analyses, and in comparison to data obtained in the literature, it may be suggested that some infection by a retrovirus occurred in a felid ancestral species in the second half of the Miocene. Therefore, the results obtained demonstrate that ERVs from Neotropical felids follow patterns which are very similar to the ones described for other mammals, also suggesting some cases of similar retrovirus lineage infecting different mammal orders.

Key words: Endogenous retrovirus, Neotropical felids, mammals, *Gammaretrovirus*, molecular evolution.

## LISTA DE ABREVIATURAS

Blast - *Basic Local Alignment Search Tool*

BI - *Bayesian Inference* (Inferência Bayesiana)

*env* - gene codificante das proteínas do envelope viral

ERV - *Endogenous Retrovirus* (Retrovírus endógeno)

FeLV- Vírus da leucemia felina

*gag* - Gene codificante das proteínas do antígeno grupo específico

*Gamma-like* - Similar ao gênero *Gammaretrovirus*

INDEL - inserções ou deleções

LINE - *Long Interspersed Nuclear Element*

LTR - *Long Terminal Repeat* (Sequência repetitiva terminal)

ML - *Maximum Likelihood* (Máxima verossimilhança)

NJ - *Neighbor-joining*

NCBI - National Center for Biotechnology Information

nr – *non-redundant nucleotide collection* (Banco de dados mantido pelo NCBI)

ORF - *Open Reading Frame* (Fase aberta de leitura)

pb - Pares de base

*pol* - Gene codificante das enzimas virais

PR - Protease

RT - Transcriptase reversa

SINE - *Short Interspersed Nuclear Element*

TEs - Elementos transponíveis

WGS - *Whole Genome Shotgun Sequencing*

XRV - *Exogenous Retrovirus* (Retrovírus exógeno)

## SUMÁRIO

<b>1. APRESENTAÇÃO .....</b>	<b>12</b>
<b>2. REVISÃO BIBLIOGRÁFICA .....</b>	<b>13</b>
<b>2.1. Retrovírus.....</b>	<b>13</b>
<b>2.2. Elementos transponíveis.....</b>	<b>14</b>
<b>2.3. Retrovírus endógenos .....</b>	<b>15</b>
2.3.1. Interação com hospedeiro .....	15
2.3.2. Taxonomia .....	18
2.3.3. Multiplicação .....	18
2.3.4. Variação genética.....	19
2.3.5. Moldando o genoma do hospedeiro.....	20
2.3.6. Retrovírus endógenos e vacinas.....	21
<b>2.4. Retrovírus e felídeos .....</b>	<b>22</b>
<b>3. OBJETIVOS .....</b>	<b>24</b>
<b>3.1. Objetivos Gerais .....</b>	<b>24</b>
<b>3.2. Objetivos Específicos.....</b>	<b>24</b>
<b>4. MATERIAIS E MÉTODOS.....</b>	<b>25</b>
<b>4.1. Amostragem .....</b>	<b>25</b>
<b>4.2. Metodologia molecular.....</b>	<b>25</b>
4.2.1. Extração e Amplificação.....	25
4.2.2. Clonagem e sequenciamento .....	26
4.2.3. Controle de contaminação .....	27
<b>4.3. Análises .....</b>	<b>27</b>
4.3.1. Processamento das sequências.....	27
4.3.2. Identificação das sequências de origem retroviral .....	28

4.3.3. Alinhamentos .....	29
4.3.4. Distância entre grupos de sequências .....	32
4.3.5. Inferência filogenética .....	32
4.3.6. Seleção .....	34
4.3.7. Recombinação.....	36
<b>5. RESULTADOS .....</b>	<b>37</b>
<b>5.1. Análise dos clones .....</b>	<b>37</b>
<b>5.2. SINEs identificados nos fragmentos sequenciados .....</b>	<b>40</b>
<b>5.3. Análises filogenéticas .....</b>	<b>44</b>
5.3.1. Alinhamentos .....	44
5.3.2. Análises intraespecíficas.....	47
5.3.3. Caracterização dos clones do grupo <i>Myotis</i> .....	60
5.3.4. Análises interespecíficas.....	68
<b>5.4. Recombinação .....</b>	<b>74</b>
<b>6. DISCUSSÃO .....</b>	<b>79</b>
<b>6.1. Diversidade e funcionalidade de ERVs no genoma dos felídeos .....</b>	<b>79</b>
<b>6.2. Relacionamento dos ERVs com seus hospedeiros .....</b>	<b>80</b>
<b>6.3. SINEs .....</b>	<b>81</b>
<b>6.4. Identificação dos retrovirus endógenos e os desafios encontrados .....</b>	<b>81</b>
<b>6.5. Linhagem <i>Myotis</i> .....</b>	<b>82</b>
<b>6.6. Relações filogenéticas .....</b>	<b>83</b>
<b>6.7. Recombinação .....</b>	<b>85</b>
<b>6.8. Evolução de ERVs através de uma visão contemporânea .....</b>	<b>85</b>
<b>7. CONCLUSÃO.....</b>	<b>87</b>
<b>8. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>88</b>
<b>9. APÊNDICES .....</b>	<b>96</b>

## **1. APRESENTAÇÃO**

Neste estudo procurou-se avaliar a diversidade de retrovírus endógenos nas espécies brasileiras de felídeos silvestres e também compreender alguns aspectos sobre a sua evolução, utilizando recursos da biologia molecular e bioinformática.

Os retrovírus endógenos constituem uma excelente ferramenta para estudos evolutivos, principalmente no que se refere à interação parasito/hospedeiro, porque as sequências desses vírus estão “fossilizadas” no genoma de seus hospedeiros. As histórias dessas interações são importantes, não apenas para a virologia, mas também para áreas como epidemiologia e medicina da conservação, porque a compreensão de eventos históricos contribui para a tomada de decisões no presente. Neste aspecto, obviamente, cada estudo sobre esse tema é um agregador de informações.

Além do exposto acima, a percepção de alguns aspectos da história evolutiva dos retrovírus endógenos suscita questionamentos sobre as consequências de se alterar, tão rapidamente, os processos de evolução patógeno/hospedeiro.

## 2. REVISÃO BIBLIOGRÁFICA

### 2.1. Retrovírus

Os retrovírus (família *Retroviridae*) consistem em um grupo diverso de vírus RNA encontrado na maioria das espécies de vertebrados. A marca diferenciada desses vírus é a sua forma de replicação, que consiste na sua transcrição reversa (RNA transcrito para DNA) e posterior integração no genoma do hospedeiro. O DNA integrado na célula do hospedeiro é denominado provírus. Dentro dos cromossomos de células infectadas, ocorre uma integração de 7-12kb de sequências provirais (Bannert *et al.*, 2010). O RNA do genoma viral é produzido pela transcrição do provírus integrado no cromossomo da célula hospedeira, utilizando a maquinaria celular de transcrição (Ravazzolo & Costa, 2007). A estratégia de integração confere ao vírus a habilidade de manter uma infecção persistente apesar da resposta imune do hospedeiro (Goff, 2001).

Os retrovírus de vertebrados são classificados em simples e complexos. Nos retrovírus simples, existem quatro principais genes dispostos na seguinte ordem: 5' *gag* - *pro* - *pol* - *env* 3' (Banner *et al.*, 2010). O gene *gag* (antígeno específico do grupo) codifica proteínas estruturais da matriz (MA), do nucleocapsídeo (NC) e do capsídeo (CA); o gene *pro*, a enzima protease (PR); o gene *pol*, as enzimas transcriptase reversa (RT) e integrase (IN); e o gene *env* codifica proteínas do envelope (subunidade de superfície (SU) e transmembrana (TM), responsáveis pela ligação e entrada do vírion na célula do hospedeiro. Nos retrovírus complexos também são encontrados genes acessórios que são usualmente necessários para ajustar diversos aspectos de sua replicação e infecção.

Em cada extremidade do genoma desses organismos encontramos as regiões não codificantes, compostas pelas regiões únicas das extremidades 5' e 3', U5 e U3, respectivamente, além da região R (repetitiva). Essas regiões, apesar de não codificantes, exercem um papel importante no ciclo de vida viral. Elas são responsáveis, por exemplo, pela ligação do t-RNA na região 5', início da transcrição reversa. Durante a transcrição reversa, as regiões U5 e U3 são duplicadas, e o DNA resultante passa a apresentar sequências idênticas (U3-R-U5), denominadas LTRs (*long terminal repeats*), nas duas extremidades (Ravazzolo & da Costa, 2007).

De acordo com o Comitê Internacional de Taxonomia Viral (ICTV) a família *Retroviridae* está subdividida em duas subfamílias: *Orthoretrovirinae*, constituída pelos gêneros *Alpha*, *Beta*, *Gamma*, *Delta*, *Epsilonretrovirus* e *Lentivirus*, e *Spumaretrovirinae*, formada pelo gênero *Spumavirus*. Esses gêneros são classificados de acordo com a homologia molecular da região da RT, a organização do genoma, a presença de genes acessórios, a morfologia (forma do capsídeo) e o local de montagem do capsídeo (membrana plasmática ou citoplasma).

## 2.2. Elementos transponíveis

Os elementos transponíveis (TEs) compõem grande parte do genoma dos eucariotos (Schaack *et al.*, 2010) e são capazes de se translocar de uma região a outra do genoma, sendo este processo denominado transposição. Os TEs podem ser divididos em duas grandes categorias segundo critérios enzimáticos e estruturais:

- retrotransposons (classe I), que utilizam um RNA intermediário, ou seja, sua transposição se caracteriza pela transcrição reversa de uma molécula de RNA, através da atividade enzimática da RT (Wicker *et al.*, 2007);

- DNA transposons (classe II), cuja transposição ocorre via DNA, através da enzima transposase.

Há várias propostas que tentam agrupar e descrever o grupo extremamente diverso de retrotransposons. Usualmente, os retrotransposons são classificados como LTRs e non-LTRs. Os retrotransposons Penelope-like e DIRS1-like são classificados à parte (Kazazian, 2004).

Os elementos que codificam proteínas necessárias para a transposição são denominados autônomos. Os retrotransposons que não possuem LTR usualmente contêm duas ORFs. ORF 1 codifica uma proteína responsável pelo transporte dos transcritos de volta para o núcleo, e a ORF2 codifica proteínas com domínios de transcriptase reversa e endonuclease (Ray & Batzer, 2011). Os retroelementos *Long Interspersed Nuclear Elements* (LINEs) fazem parte desta categoria (Symer & Boeke, 2010). Os retrotransposons que possuem LTR apresentam ORFs para as proteínas Gag e Pol, enquanto que os retrovírus endógenos possuem também a ORF para a proteína Env. Desse modo, os retrovírus endógenos podem formar partículas infecciosas.

Existem retrotransposons que não codificam as proteínas necessárias para a transposição, que são denominados não autônomos. É a esta categoria que pertencem os *Short Interspersed Nuclear Elements* (SINEs), os quais, acredita-se sejam proliferados com auxílio de LINEs (Symer & Boeke, 2010).

Em estudos realizados com *Monodelphis domestica*, investigadores descobriram que grande parte das inovações evolutivas que distinguem os Metatheria dos Eutheria não foram causadas pela diferenciação em regiões codificantes, mas sim em regiões não codificantes, e que os TEs podem ser considerados a principal força inovadora na evolução dos mamíferos (Ray & Batzer, 2011).

## **2.3. Retrovírus endógenos**

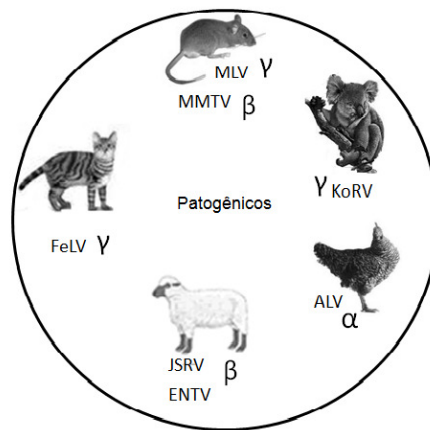
### **2.3.1. Interação com hospedeiro**

Um retrovírus endógeno (ERV) surge quando o DNA viral se integra a uma célula germinativa, passa a fazer parte do patrimônio genético do hospedeiro e é transmitido através das gerações de maneira tipicamente mendeliana (Weiss, 2006). No decorrer da evolução, as sequências destes retrovírus acumulam mutações e inserções/deleções, as quais, via de regra, os tornam defectivos (Gifford *et al.*, 2005). Devido à sua disseminação dentro do hospedeiro, as novas inserções podem causar câncer e doenças de origem genética (Goff, 2001). Entretanto, apesar do grande potencial deletério, a maioria delas tem caráter neutro, sem qualquer impacto funcional no organismo. Isso ocorre porque existem mecanismos, como a metilação, que reduzem seu potencial (Goff, 2001; Maksakova *et al.*, 2008). Além disso, as sequências LTRs dos ERVs são promotoras mais fracas se comparadas com as dos retrovírus exógenos (XRVs), refletindo uma seleção contra a patogenicidade (Goff, 2001).

Apesar da maioria dos ERVs ser defectiva, isso não significa que eles não apresentem potencial para patogenicidade. Um exemplo clássico ocorre com o vírus da leucemia felina (FeLV), cuja variante viral FeLV-B surgiu da recombinação entre subgrupo FeLV-A e o FeLV endógeno (enFeLV) (Overbaugh & Bangham, 2001). A recombinação pode alterar a especificidade entre hospedeiros e propriedades de replicação do vírus. Portanto, sempre que surgirem novas variantes de um retrovírus exógeno, a contribuição de um endógeno deveria ser avaliada (Goff, 2001). Além do



FeLV, há ERVs co-existindo com uma forma exógena que lhe é correspondente, em camundongos (*mouse mammary tumor virus* (MMTV) e *murine leukaemia virus* (MLV)), em ovinos e caprinos (*Jaagsiekte sheep retrovirus* (JSRV), *enzootic nasal tumor virus* (ENTV)), em suínos (*porcine ERV*), em alguns Galliformes (*avian leukaemia virus* (ALV)) e algumas populações de Koalas (Koala ERV (KoRV)) (Denner, 2010). A maioria dessas formas exógenas está associada diretamente a doenças (Figura 1).



**Figura 1.** Formas exógenas de retrovírus associadas diretamente a doenças, que possuem um ERV muito similar coexistente no hospedeiro. As letras gregas representam o gênero ao qual pertence o retrovírus: *Alpharetrovirus* ( $\alpha$ ), *Betaretrovirus* ( $\beta$ ) e *Gammaretrovirus* ( $\gamma$ ).

Algumas inserções de ERVs podem ser vantajosas porque protegem o hospedeiro de superinfecções, geralmente provocadas pelo retrovírus exógeno correspondente (Löwer *et al.*, 1996; Arnaud *et al.*, 2007; Pontius *et al.*, 2007). O retrovírus endógeno enJRSV (*Jaagsiekte sheep retrovirus*) tem um importante papel na evolução da ovelha doméstica, porque ele bloqueia a replicação do exógeno JRSV, além de participar no desenvolvimento da placenta (Denner, 2010). No homem, a proteína syncytin-1, produto do gene *env* do HERV-W desempenha papel importante na formação da placenta, demonstrando a sua relevância na reprodução (Jern & Coffin 2008). Os vertebrados superiores possuem muitas cópias de retrovírus endógenos integradas ao seu genoma (Herniou *et al.*, 1998; Gifford *et al.*, 2005) - 8% do genoma humano (Weiss, 2006) e aproximadamente 4% do genoma do gato doméstico (Pontius *et al.* 2007). Poucas inserções de ERVs são recentes (e.g., Tarlinton *et al.*, 2006; Arnaud *et al.*, 2007), sendo que a grande maioria destas cópias são antigas, mantendo uma relação

filogenética distante com os exógenos existentes. Uma possível explicação para este fato seria que, durante a evolução, a inserção de um retrovírus em uma célula germinativa pode ter sido vantajosa para a progênie que a continha, propiciando resistência a superinfecções provocadas pela sua forma exógena. Ao longo do tempo, isso teria favorecido a fixação da forma endógena e a extinção da exógena (Löwer *et al.*, 1996).

A grande maioria dos indivíduos de uma população compartilha sequências de ERVs muito similares (Arnaud *et al.*, 2007). Isso se deve ao fato de que o aparecimento de um retrovírus endógeno em uma linhagem (população) filogeneticamente relacionada, geralmente é o resultado de um simples evento de integração no genoma de um hospedeiro (Chessa *et al.*, 2009). A proporção de ERVs fixados dentro de uma população irá depender de vários fatores, tais como: tempo decorrido desde a inserção, a capacidade de disseminação do vírus, flutuação do tamanho efetivo populacional e pressão de seleção (Gifford & Tristem, 2003). Entre linhagens (por exemplo, espécies, subespécies), encontram-se diferentes padrões de diversidade, o que sugere um rearranjo desta ao longo do tempo (milhões de anos) (Goff, 2001).

Devido aos registros “fósseis” encontrados no genoma, é possível reconstituir a história evolutiva dos retrovírus numa escala de milhões de anos. Vários estudos demonstraram a evolução dos ERVs ao longo do tempo e, adicionalmente, contribuíram para o entendimento da evolução de seus hospedeiros (Arnaud *et al.*, 2007; Katzourakis *et al.*, 2007; Romano *et al.*, 2007; Gifford *et al.*, 2008). Estudos de ERVs em primatas, roedores e outras espécies indicam um longo relacionamento entre retrovírus e vertebrados, sendo algumas inserções datadas em vários milhões de anos (Romano *et al.*, 2007; Keckesova, 2009).

Uma longa história de associação entre “parasito” e hospedeiro frequentemente leva à coevolução e coespeciação, de modo que a filogenia desses dois organismos tende a ser concordante (Page & Holmes, 1998; Poss *et al.*, 2002). Assim, a comparação entre as filogenias do parasito e seus hospedeiros fornece um registro indireto dos eventos de especiação e pode ajudar a inferir como eles se associam (Huyse *et al.*, 2005).

Em vírus que causam infecções persistentes, como os retrovírus, os padrões de distribuição são geralmente moldados por vicariância, e, sendo assim, uma coespeciação

é um resultado previsível (Holmes, 2004). Entretanto, a mudança de hospedeiro também tem moldado fortemente a diversidade de patógenos através do tempo (Woolhouse *et al.*, 2005) e é um dos fatores mais importantes para o surgimento de doenças emergentes que transpõem a barreira de espécie, como a AIDS.

### **2.3.2. Taxonomia**

A categoria mais importante na classificação dos vírus é a família (Comitê Internacional de Taxonomia Viral — ICTV). Diferentemente de organismos celulares (Archaea, Bacteria e eucariotos), que podem ser comparados em uma mesma filogenia, por compartilharem genes homólogos, as relações evolutivas nos vírus são estabelecidas dentro das famílias e muito dificilmente entre elas. Geralmente, vírus classificados como pertencentes a uma mesma família apresentam estrutura genômica similar e genes homólogos.

ERVs possuem homologia molecular com XRVs, especialmente através do gene polimerase. Portanto, existe uma tendência de agrupá-los em classes conforme sua similaridade com XRVs: (i) classe I, ERVs filogeneticamente relacionados com XRVs *Gamma* e *Epsilonretrovirus*; (ii) classe II, ERVs relacionados com XRVs *Lentivirus*, *Alpha*, *Beta* e *Deltaretrovirus*; e (iii) classe III, ERVs relacionados com XRVs *Spumavirus* (Gifford *et al.*, 2005; Blomberg *et al.*, 2009). Pela própria natureza dessa classificação, a reconstrução filogenética e classificação são inferidas através de propriedades que evoluíram antes da endogenização.

Existem controvérsias em relação à nomenclatura dos ERVs, especialmente dos de origem humana (HERVs). Esses retrovírus são classificados em “famílias” de acordo com o t-RNA utilizado pelo vírus. Alguns autores sugerem que a sistemática e a nomenclatura dos ERVs sejam reformuladas e unificadas (Blomberg *et al.*, 2009; Mayer *et al.*, 2011).

### **2.3.3. Multiplicação**

Um ERV coloniza o genoma de um hospedeiro de duas maneiras distintas, o que influencia diretamente o seu número de cópias e seu modo de evolução. Durante a replicação, o DNA proviral é transcrito para mRNA pela maquinaria celular de transcrição e, a transcriptase reversa converte as moléculas de mRNA em DNA fita dupla (provírus) que é integrado ao genoma do hospedeiro (retrotransposição). Ou, de

modo similar aos XRVs, duas cópias de RNA viral (dois genomas) são empacotadas dentro de um vírion que pode infectar outras células (reinfecção). Quando a célula está infectada por dois retrovírus muito similares é possível gerar um vírus heterozigoto que carrega dois genomas diferentes (Lewin, 2008).

Portanto, novas inserções vão surgindo via retrotransposição (intracelular) ou via reinfecção (extracelular). De um modo geral, a disseminação via reinfecção gera menor número de cópias (Löwer *et al.*, 1996).

A retrotransposição pode seguir duas rotas: em *cis*, quando o retroelemento codifica suas proteínas necessárias para a transposição; ou complementação em *trans* onde as proteínas são fornecidas por outro retrovírus (endógeno ou exógeno) que está infectando a mesma célula. A retrotransposição em *cis* não requer que o gene *env* (envelope) seja funcional; já a complementação em *trans* não requer nenhum gene funcional, simplesmente exige que um promotor e outros *motifs* estejam intactos (Belshaw *et al.*, 2004).

O ciclo de vida de um retrovírus é um grande gerador de diversidade devido à alta taxa de mutação e recombinação que pode ser atingida durante a replicação. Cada evento de invasão de um retrovírus no genoma representa uma nova linhagem, como é exemplificado por Polani *et al.* (2010) em um estudo com FeLV endógeno. Isso ocorre porque uma linhagem não é uma simples cópia de um retrovírus, mas um conjunto de cópias interligadas pelos mesmos eventos históricos. As características das linhagens atuais de ERVs são influenciadas por vários fatores, tais como: a idade e atividade da retrotransposição, efeitos demográficos do hospedeiro e pressão de seleção (Grifford & Tristem, 2003). A taxa de amplificação da população viral tende a ser muito maior no início da infecção e a declinar com o passar do tempo graças à ação de mecanismos de defesa do hospedeiro. Manter um genoma funcional ao longo do tempo, provavelmente requer replicação e, para isso, o genoma viral poderia ser reciclado via recombinação (Weiss, 2006).

#### **2.3.4. Variação genética**

Sequências de ERVs são difíceis de comparar porque divergem a uma taxa semelhante à de um pseudogene (Blomberg *et al.*, 2009). Como os ERVs são, em sua maioria, integrações antigas (Gifford & Tristem, 2003) geralmente ocorre uma alta

divergência entre as sequências podendo causar incertezas no alinhamento e, conseqüentemente, na filogenia. Quando as sequências são muito divergentes, com menos de 30% de identidade em sequências de proteína (*twilight zone* de um alinhamento), a acurácia do alinhamento diminui consideravelmente (Do *et al.*, 2005). Um alinhamento de nucleotídeos é confiável quando apresenta 50% ou mais de identidade. Assim, a ferramenta primária utilizada para a classificação dos ERVs é o gene *pol*, que é evolutivamente mais conservado (Blomberg *et al.*, 2009).

Outros eventos que dificultam o alinhamento são as recombinações e inserções. A título de exemplo, aproximadamente 20% do grupo HERV-K (HML2) sofreu rearranjos em seu genoma desde sua integração, inclusive por meio de recombinações com diferentes provírus (Griford & Tristem, 2003). É certo que, quando indivíduos trocam informação genética por recombinação, eles também compartilham parte de suas histórias evolutivas (Leyme & Posada, 2009).

O grau de variação pós-integracional pode ser estimado pelo grau de nucleotídeos não idênticos na região da LTR, pelo número de códons de parada, pelas mutações que resultam na mudança da fase de leitura no gene *pol*, pela divergência da sequência consenso de um grupo de sequências similares (Blomberg *et al.*, 2009).

### **2.3.5. Moldando o genoma do hospedeiro**

Na metade do século passado, Barbara McClintock questionou a estabilidade do genoma, conceito proposto naquela época, através da identificação de elementos transponíveis no genoma do milho (McClintock 1950). A partir do marco histórico estabelecido pelas descobertas de McClintock, vários outros elementos móveis foram descritos em uma variedade de organismos, inclusive nos vertebrados. O genoma dos vertebrados alberga vários tipos de sequências que se movem pelo genoma, dentre elas, a dos retrovírus endógenos.

Existem várias maneiras pelas quais um ERV pode influenciar a evolução de um genoma, e uma das mais impactantes é a transcrição reversa, com integração da sequência viral no genoma do hospedeiro. Durante essa fase do ciclo viral, é possível que ocorram vários eventos: perda da função de um gene, causada pela inserção do DNA viral dentro deste, alteração na expressão de genes devido à inserção de LTRs em sua proximidade e ativação de oncogenes ou inativação de genes supressores de tumores

(Denner, 2010). É possível também que surjam novidades evolutivas no genoma do hospedeiro, como o clássico exemplo da expressão da amilase salivar em humanos provocada pela integração do HERV-E (Jern & Coffin, 2008). Outro evento é a criação de pseudogenes processados quando um RNA maduro (mRNA) sofre transcrição reversa e é incorporado no genoma do hospedeiro. Além disso, sequências celulares podem se recombinar com sequências retrovirais e, em seguida, ser transportadas juntamente com o retrovírus e inseridas em um novo local do genoma do hospedeiro (Lewin, 2008). Inserções de sequências provirais também podem influenciar nos processos de rearranjo dos cromossomos, através da recombinação homóloga entre provírus localizados em diferentes regiões do genoma. Todos esses processos acima mencionados contribuem para a plasticidade do genoma.

Os genes do Complexo Principal de Histocompatibilidade (MHC) possuem ERVs integrados em introns e estas inserções contribuem para a plasticidade de seus genes, resultando em uma grande variedade de alelos (Doxiadis *et al.*, 2008). A plasticidade do MHC tem um importante papel na evolução dos vertebrados (Loisel *et al.*, 2008).

A constante batalha entre hospedeiros e retrovírus também contribuiu para a diversidade do genoma do hospedeiro. O hospedeiro, no curso de sua evolução, adquiriu genes que codificam fatores de restrição retrovirais que bloqueiam a infecção de alguns retrovírus. Um exemplo desses fatores é o APOBEC3, um complexo de proteínas pertencentes à família citidina deaminase, que causam hipermutações no genoma viral (Harris & Lidammet, 2004).

Assim, como os retrovírus estão intrinsecamente relacionados com o genoma de seus hospedeiros, durante o curso de sua evolução, um ERV poderá ser parasito, simbiote ou comportar-se de maneira neutra (Blikstad *et al.*, 2008).

### **2.3.6. Retrovírus endógenos e vacinas**

Muitas vacinas de vírus vivo atenuado para animais são produzidas utilizando cultura de células provenientes de camundongos, suínos, galinhas e gatos. Estes animais produzem retrovírus endógenos com replicação ativa (Miyazawa, 2010).

Recentemente, Miyazawa *et al.* (2010) descobriu que certas vacinas vivas atenuadas para cães estavam contaminadas com RD114 infeccioso. Vacinas apresentando risco de contaminação com RD114, também poderiam ter sido utilizadas

em gatos domésticos e silvestres, furões e nos grandes felinos (Miyazawa, 2010). A maioria destas espécies, à semelhança dos cães, não possui o provírus RD114 integrado no seu genoma. Um recente estudo experimental demonstrou que o RD114 não causa efeito patogênico algum em cães (Narushima *et al.*, 2011). Além do mais, efeitos patogênicos não foram reportados até o momento, considerando-se o grande número de cães vacinados anualmente. Entretanto, segundo Dewannieux *et al.* (2010), os riscos de se utilizar vacinas produzidas a partir de linhagens celulares que produzem partículas virais de ERVs não podem ser ignorados.

#### **2.4. Retrovírus e felídeos**

Recentemente, os felídeos têm sido foco de vários estudos com retrovírus exógenos, principalmente com o vírus da imunodeficiência felina (FIV) (Biek *et al.*, 2003; Poss *et al.*, 2007; Franklin *et al.*, 2007). Esse vírus apresenta características genéticas e manifestações clínicas similares à AIDS, sendo considerado um modelo de estudo para essa doença (O'Brien *et al.*, 2006; Tandon *et al.*, 2008).

Existem 37 espécies de felídeos, sendo que dez ocorrem na América do Sul e oito destas são encontradas no Brasil. Infelizmente, essas espécies estão quase todas listadas como ameaçadas ou quase ameaçadas na lista vermelha mundial ([www.iucnredlist.org](http://www.iucnredlist.org)) e, em decorrência disso, parte da diversidade dos ERVs pode estar sendo perdida junto com essas espécies e populações.

Os felídeos modernos são representados por espécies bem distribuídas, que apresentam um bom registro fóssil e muitos dados moleculares, o que possibilitou inferir hipóteses sobre a sua origem e diversificação (Johnson *et al.*, 2006). Estudos recentes indicam que os felídeos modernos iniciaram a sua diversificação na segunda metade do Mioceno (há aproximadamente 10 milhões de anos) e que a migração moldou grande parte dessa diversidade (Johnson *et al.*, 2006). A colonização dos felídeos na América do Sul ocorreu após o fechamento do istmo do Panamá, no início do Plioceno (há aproximadamente 3,5 milhões de anos) (Indrusiak & Eizirik, 2003; Johnson *et al.*, 2006). Segundo Johnson *et al.* (2006), as espécies neotropicais são divididas em três linhagens: Jaguatirica (composta pelo *Leopardus pardalis*, *L. wiedii*, *L. jacobita*, *L. colocolo*, *L. geoffroyi*, *L. guigna* e *L. tigrinus*), Puma (*Puma concolor* e

*P. yagouaroundi*) e *Panthera (Panthera onca)*. Alguns felídeos neotropicais são amplamente distribuídos, como, por exemplo, *P. concolor*. Enquanto outros, como *L. geoffroyi* e *L. guigna*, possuem uma distribuição um pouco mais restrita. Algumas espécies são simpátricas, enquanto outras apresentam zonas de contato, sendo que algumas áreas comportam inclusive indivíduos híbridos de *L. colocolo* e *L. tigrinus* (Johnson *et al.*, 1999), e de *L. geoffroyi* e *L. tigrinus* (Trigo *et al.*, 2008). A grande maioria possui hábitos terrestres, demonstrando certa plasticidade com relação ao habitat. No entanto, algumas linhagens fogem desse padrão; um exemplo disso é *L. wedii*, que é mais arborícola e vive preferencialmente em áreas florestadas (Mattern & McLennan, 2000; Indrusiak & Eizirik, 2003).

O conhecimento de retrovírus endógenos nos felídeos está praticamente restrito ao gato doméstico. Um dos ERVs mais conhecidos é o RD114. Estudos indicam que o RD114 é um retrovírus recombinante entre *Felis catus* ERV (FcEV) e *Baboon endogenous virus* (BaEV), proveniente do babuíno, sugerindo uma transmissão interespecies muito antiga (Van Der Kuyl *et al.*, 1995). Outro ERV bastante estudado é a forma endógena do vírus da leucemia felina, que é creditado como sendo de origem murina. A forma exógena do FeLV, como descrito anteriormente, ainda circula nas populações e é patogênica (Tandon *et al.*, 2008). Uma análise filogenética utilizando sequências LTR provenientes de enFeLV de um gato da raça abissínia, demonstrou a existência de duas linhagens, sugerindo que pelo menos duas diferentes infecções ocorreram na história do gato doméstico (Pontius *et al.*, 2007). RD114 e enFeLV são restritos ao gênero *Felis* corroborando a hipótese filogenética sugerida para os felídeos modernos (Johnson *et al.*, 2006).

Recentemente, novos ERVs foram identificados no genoma do gato doméstico. Pontius *et al.* (2007) descreveram cinco novas linhagens (FERV1-5), e Yuhki *et al.* (2008) localizaram, na região do MHC classe I, os ERVs FERVmlu1 e FERVmlu2, que apresentam homologia com um ERV proveniente do morcego *Myotis lucifugus*.



### **3. OBJETIVOS**

#### **3.1. Objetivos Gerais**

O objetivo deste estudo foi a caracterização molecular de retrovírus endógenos em felídeos neotropicais, avaliar sua distribuição nas oito espécies estudadas e fazer inferências sobre alguns aspectos de sua evolução.

#### **3.2. Objetivos Específicos**

1. Identificar sequências de ERVs em felídeos neotropicais com uma análise detalhada dos diferentes elementos encontrados.
2. Estudo dos relacionamentos filogenéticos dos ERVs nas oito espécies avaliadas, contribuindo para uma melhor compreensão de alguns aspectos de sua evolução, através da identificação *in silico*, de sequências similares em outros organismos.
3. Verificar a existência de um padrão coevolutivo nos ERVs avaliados.
4. Gerar subsídios para estudos referentes à evolução de parasito/hospedeiro.

## 4. MATERIAIS E MÉTODOS

### 4.1. Amostragem

Amostras de sangue ou tecido muscular de gato-do-mato-pequeno (*Leopardus tigrinus*), gato-maracajá (*Leopardus wiedii*), jaguatirica (*Leopardus pardalis*), gato-palheiro (*Leopardus colocolo*), suçuarana (*Puma concolor*) e de onça-pintada (*Panthera onca*) foram obtidas no Laboratório de Biologia Genômica e Molecular, Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Uma amostra de tecido muscular de gato-do-mato-grande (*Leopardus geoffroyi*-MCP1781) foi obtida do Museu de Ciências e Tecnologia da PUCRS, e uma amostra de sangue gato-mourisco (*Puma yagouarondi*) foi obtida de um animal atropelado que recebeu atendimento clínico no Hospital de Clínicas Veterinárias da Universidade Federal do Rio Grande do Sul, em 03 de outubro de 2008. Foi estudado um indivíduo por espécie e pelo menos 25 clones de cada indivíduo.

### 4.2. Metodologia molecular

#### 4.2.1. Extração e Amplificação

O DNA genômico foi extraído utilizando-se o PureLink Genomic DNA Kit (Invitrogen), seguindo-se as recomendações do fabricante. Foram estudados fragmentos compostos por parte do gene da protease e parte do gene da polimerase (região RT), os quais demonstraram ser conservados em estudos anteriores (Herniou *et al.*, 1998; Jern *et al.*, 2005; Xiao *et al.*, 2008), possibilitando reconstruir hipóteses filogenéticas robustas. Para isso, foram utilizados quatro primers degenerados, conforme Tristem (1996). Esses primers foram desenhados com base em seqüências conservadas nos sítios ativos da protease (DTGA) e da transcriptase reversa (YMDD / YVDD).

Os primers são os seguintes:

Protease (primer sense):

(PRO) 5'GTT/GTTIG/TTIGAT/CACIGGIG/TC3';

Transcriptase Reversa (primer antisense):

(CT) 5' AGIAGGTCA/GTCIACA/GTAC/GTG 3',

(JO) 5' ATIAGIAG/TA/GTCA/GTCIACA/GTA3',

(EM) 5' ATIAGIAG/TA/GTCA/GTCCATA/GTA 3'; onde I é inosina.

As condições das reações em cadeia da polimerase (PCR) foram as seguintes: desnaturação a 80° C por 2 min; seguida por 35 ciclos a 45°C por 30 seg para anelamento, 74°C por 60 seg para extensão e 94°C por 30s para desnaturação; e um ciclo final com temperaturas de 45°C por 3 min e 72°C por 10 min.

Para cada par de primers, utilizaram-se 100 ng de DNA genômico, 2 U de *Taq* polimerase, 40 pmol de cada primer, 200 µM de desoxinucleotídeos trifosfato e tampão 1× da PCR (10 mM Tris [pH 8.3], 50 mM KCl, 1.5 mM MgCl<sub>2</sub>) para um volume final de 25µl.

#### **4.2.2. Clonagem e sequenciamento**

A separação dos fragmentos de DNA amplificados foi realizada através de eletroforese em gel agarose de baixo ponto de fusão a 1% em tampão TAE (TRIS-acetato 0,04M e EDTA 0,001M), corado com brometo de etídio (Sambrook *et al.*, 1989). Produtos entre 600 e 1200 pb foram excisados (Herniou *et al.*, 1998) e purificados utilizando o kit Wizard SV gel (Promega). A concentração de DNA foi estimada por comparação com o marcador de peso molecular MassRuller (Fermentas), com auxílio do programa Image J v.1.41 (<http://rsb.info.nih.gov>). Os fragmentos foram clonados utilizando-se o vetor pGEM-T Easy Vector e a cepa de *E. coli* competente JM 109 (Promega), de acordo com as recomendações do fabricante. O método utilizado para transformação foi choque térmico (Sambrook *et al.*, 1989). O DNA plasmídeo foi obtido através do método de lise alcalina (Sambrook *et al.*, 1989) ou kit comercial (Promega) e digerido com a enzima EcoRI a fim de verificar o tamanho dos fragmentos clonados. O sequenciamento dos clones foi realizado diretamente dos plasmídeos purificados. Ambas as fitas foram sequenciadas utilizando-se os *primers* universais: M13 (senso e antissenso) ou T7 e SP6. As amostras foram sequenciadas utilizando-se o *kit* BigDye Terminator v. 3. 1 (Applied Biosystems), de acordo com as instruções do fabricante. Os produtos foram analisados utilizando o sequenciador automático ABI-PRISM 3100 Genetic Analyzer (Applied Biosystems).

### 4.2.3. Controle de contaminação

Durante os procedimentos laboratoriais, existem riscos de contaminação das amostras por amplicons/plasmídeos gerados em um experimento anterior, o que se caracteriza como um acidente comum em biologia molecular, contemporâneo na retrovirologia (Hué *et al.*, 2010). Para evitar essa contaminação, alguns cuidados foram tomados:

1. Toda a manipulação do experimento laboratorial, incluindo preparo de reagentes e lavagem de material foi realizada apenas pela autora deste trabalho, com objetivo de minimizar os erros.
2. A extração do DNA genômico das amostras e preparação da reação da PCR foram realizadas em salas distintas das salas de clonagem e purificação de plasmídeos.
3. O material de cada espécie de felídeo foi manipulado isoladamente, desde a sua extração até o sequenciamento. Desse modo, o início da manipulação de uma nova espécie ocorria somente após a finalização dos procedimentos com a espécie anterior.
4. A reação da PCR foi cuidadosamente avaliada com controles negativos.

## 4.3. Análises

### 4.3.1. Processamento das sequências

Os cromatogramas foram processados com os programas Pregap4 e Gap4 (<http://staden.sourceforge.net>). Para cada clone foi gerada uma sequência consenso e foram removidas tanto as sequências de bases correspondentes ao vetor, em ambas as fitas, quanto às bases de baixa qualidade localizadas nas extremidades dos fragmentos (escore de qualidade Phred < 21).

As sequências obtidas foram nomeadas da seguinte forma: as letras iniciais referem-se à espécie de felídeo analisada, as duas letras seguintes representam o *primer* (referente à RT) do fragmento amplificado, em seguida, está indicado o tamanho do fragmento em pares de base (pb) e, por fim, o número do clone. Exemplo: Py (*Puma yagouarondi*) JO (*primer*) 1000 (tamanho) 1 (número do clone).

### 4.3.2. Identificação das sequências de origem retroviral

As sequências de nucleotídeos similares a retrovírus foram identificadas por comparação com sequências disponíveis no NCBI (<http://www.ncbi.nlm.nih.gov>) utilizando-se o programa BLAST (Basic Local Alignment Search Tool) (Altschul *et. al.* 1997). Resumidamente, o programa BLAST procura sequências através de uma busca heurística baseada em palavras (nucleotídeos ou proteínas) comuns a uma sequência *query* (sequência de entrada) e uma *subject* (depositada no banco de dados) (Pertsemlidis & Fondon 2001). Foram utilizadas as modalidades BLASTn (nucleotídeos vs. nucleotídeos) e BLASTx (nucleotídeos vs. proteína) do programa BLAST. Além disso, a modalidade BLASTp (proteína vs. proteína) foi utilizada para verificar a presença dos domínios similares à transcriptase reversa e protease de ORFs parciais identificadas no programa ARTEMIS 11.22 (<http://www.sanger.ac.uk/Software/Artemis>). Para melhor caracterizar determinadas sequências e/ou linhagens, utilizou-se, em alguns momentos, a modalidade tBLASTn (proteína vs. nucleotídeo traduzido), principalmente quando se pretendia comparar sequências pertencentes a organismos evolutivamente distintos. No início deste estudo procurou-se aperfeiçoar as buscas, modificando alguns parâmetros no programa BLAST, tais como: a penalidade para *gaps* (descontinuidade das palavras) e matriz substituição de aminoácidos. Entretanto, verificou-se que os resultados, em geral, não eram diferentes daqueles obtidos utilizando-se o padrão do programa. Razão pela qual, foi utilizado o padrão do programa nas análises. Em todas as buscas somente foram considerados como clones positivos aqueles que possuíam mais de 100pb similares a retrovírus e com valor esperado (*e value*) menor que 0.05. Esse valor reflete o número de falsos positivos, ou seja, a probabilidade do alinhamento ter ocorrido ao acaso (Bottu, 2009).

Para as buscas, utilizou-se o banco de dados nr, que é composto por todas as sequências do GenBank (mantido pelo NCBI), EMBL (European Molecular Biology Laboratory), DDBJ (DNA DATA BANK of Japan) e PDB (Protein Data Bank). Para algumas análises mais específicas, foram realizadas buscas de sequências no banco de dados WGS (*Whole Genome Shotgun Sequencing*) do GenBank, que recebe sequências de projetos de montagem de genomas.

Com o propósito de melhor caracterizar um restrito número de sequências de ERVS, que demonstraram diferenças no alinhamento em relação ao tamanho e/ou regiões não alinhadas, foi realizada uma busca de elementos repetitivos, utilizando-se o programa RepeatMasker (<http://www.repeatmasker.org>). Em situações isoladas utilizou-se a ferramenta de busca BLAT search (Blast-Like Alignment Tool), disponível no *site* <http://genome.ucsc.edu/cgi-bin/hgBlat>.

Todas as buscas de dados foram realizadas no período de agosto a dezembro de 2011.

### **4.3.3. Alinhamentos**

#### **Sequências provenientes do GenBank**

Uma lista contendo sequências provenientes do GenBank utilizadas nos diferentes alinhamentos encontra-se no Apêndice A. O critério geral para a escolha das sequências foi o seguinte: 1) a similaridade com clones obtidos neste estudo com base no programa BLAST e 2) a literatura, principalmente, a relacionada com o grupo de sequências estudado.

#### **Alinhamento das sequências**

##### **A. Nucleotídeos**

Em primeiro lugar, realizou-se um alinhamento de nucleotídeos que foi construído utilizando-se o programa DIALIGN 2.2.1 (Morgenstern, 1999) disponível no pacote de programas do JEmboss 1.5 (<http://emboss.open-bio.org>). Nesse programa, o alinhamento global não é forçado, permitindo alinhar sequências altamente variáveis. Isso é importante em situações nas quais a maioria das sequências apresenta similaridades locais separadas por regiões não relacionadas (Lemey & Higgins, 2009). Apenas as sequências com similaridade à *Gammaretrovirus* foram incluídas no alinhamento, pois sequências relacionadas com outros grupos apresentavam-se muito divergentes e a acurácia do alinhamento poderia diminuir consideravelmente. (Do *et al.*, 2005).

Posteriormente, as sequências foram agrupadas conforme suas similaridades para facilitar a visualização dos sítios homólogos. As sequências que apresentaram porções muito divergentes foram comparadas (par a par) com as que apresentavam similaridade, para que se localizassem fragmentos heterólogos. Em alguns casos, essa comparação foi

feita com sequências obtidas no GenBank. A realização desse procedimento se deu por meio de um alinhamento local feito com o programa WATER, do pacote JEmboss 1.5. Considerou-se o valor 10 para penalidade de abertura do *gap* e 0.5 para a sua extensão; ambos são valores do *default* do programa. Também foi utilizado o programa BLASTn, limitando a busca aos táxons agrupados na família *Retroviridae*; isto foi feito para minimizar o retorno de ruído nas buscas, recuperando, assim, apenas sequências retrovirais.

Através dessas duas estratégias, foi possível mapear quais sequências apresentavam fragmentos heterólogos, delimitar o local onde eles estavam inseridos e identificá-los. Em seguida, esses fragmentos foram removidos das sequências.

O procedimento descrito acima foi um processo inicial, exploratório. Neste estudo, realizaram-se diferentes análises cujos alinhamentos foram construídos com diferentes conjuntos de sequências. Esses alinhamentos estão descritos com mais detalhes nos resultados.

## **B. Proteínas**

### **B1. Alinhamento interespecífico**

Inicialmente, as sequências de aminoácidos dos clones obtidas neste estudo foram deduzidas através do programa Artemis. Aquelas sequências que apresentavam fase de leitura correta e sem códons de terminação prematuros foram alinhadas com o programa Clustal W, incluído no programa MEGA 5 (Tamura *et al.*, 2011). Esse alinhamento, contendo os clones similares à *Gammaretrovirus* e uma sequência de FeLV proveniente do GenBank (Apêndice A), foi utilizado como base para localizar *indels* (inserções ou deleções) responsáveis pela mudança da fase de leitura quando as demais sequências foram adicionadas. Portanto, nesta fase, trabalhou-se simultaneamente com aminoácidos e nucleotídeos. O procedimento acima foi adotado porque muitas sequências apresentavam-se degeneradas. A manipulação e edição das sequências foi realizada através do programa MEGA 5. Em seguida, foram acrescentadas sequências provenientes do GenBank, segundo a similaridade obtida com o programa BLAST. A esse acréscimo, seguiram-se um refinamento do alinhamento de aminoácidos, realizado por meio do programa DIALIGN2 e, uma edição manual no MEGA 5. Esse alinhamento foi utilizado como base em todas as análises realizadas com aminoácidos e, dependendo da análise, sequências foram acrescentadas ou retiradas.

A partir desse alinhamento inicial, também foram construídos dois alinhamentos parciais cujos fragmentos dos genes da protease ou da transcriptase reversa foram analisados separadamente. Este procedimento teve como objetivo realizar uma avaliação individual de cada gene.

Por fim, construiu-se um alinhamento que englobava um grande número de sequências partindo-se apenas de um fragmento da transcriptase reversa, por ser mais conservado. Nessa matriz, também foram adicionadas sequências de aminoácidos de retrovírus exógenos (representantes de cada gênero) e algumas sequências de ERVS descritos na literatura (Apêndice A).

Para análise filogenética, os alinhamentos foram convertidos em formato *nexus* e *phylip*. Maiores detalhes dos alinhamentos utilizados nas análises são descritos nos resultados.

## **B2. Alinhamento intraespecífico**

Depois de realizado um alinhamento interespecífico com aminoácidos (conforme apresentado na Seção B1), as sequências de aminoácidos foram separadas por espécie, em alinhamentos individualizados.

Adotou-se esta estratégia porque as sequências dos clones obtidos neste estudo foram mais similares entre grupos de retrovírus do que dentro das espécies de felídeos. Em alguns casos, foram incluídas mais sequências (similares ou que possuíam fragmento curto) com objetivo de maximizar o número de sequências. Um refinamento no alinhamento de aminoácidos foi realizado no programa DIALIGN2. Da mesma forma que no alinhamento geral, uma região variável entre a região da protease e transcriptase reversa foi removida, conforme descrito nos resultados.

O objetivo do alinhamento individualizado foi descrever o padrão de diversidade dentro das espécies. Devido à alta divergência de algumas sequências, conforme descrito acima, os clones de ERVs não pertencentes à *Gammaretrovirus* foram analisados separadamente.

## **Verificando a saturação**

Para uma reconstrução filogenética acurada, as sequências devem conter variações, mas estas não podem ser tão divergentes a ponto de acumular substituições múltiplas e, conseqüentemente, deixar de refletir um relacionamento filogenético correto (Xia, 2009). Comumente, o número de transições é maior que o de transversões,



mas esta proporção pode se alterar em função da saturação. A saturação pode ser definida como informações históricas (substituições) reescritas várias vezes em sequências de DNA. Uma representação visual do grau de saturação em um alinhamento pode ser obtida através da plotagem do número de transições e transversões observadas *versus* a distância genética corrigida por um modelo de evolução. Quando um gráfico representando as distribuições de transições e transversões apresenta o padrão de uma reta, é um indicativo que não há saturação no alinhamento. Neste estudo, essa ferramenta foi utilizada através do programa Dambé (Xia & Xie, 2001) para verificar o grau de saturação nos alinhamentos com sequências de nucleotídeos empregados nas reconstruções filogenéticas.

#### **4.3.4. Distância entre grupos de sequências**

Grupos de sequências homólogas foram identificados nas filogenias e, posteriormente, foi calculada a distância genética (distância p) entre os grupos. A distância p é a proporção de sítios homólogos diferentes entre as sequências. Assim, ela representa a distância observada, sem correções. Nas análises, foi utilizado o programa MEGA 5.

#### **4.3.5. Inferência filogenética**

A reconstrução de filogenias inferidas através do alinhamento de aminoácidos e nucleotídeos foi baseada nos seguintes métodos: máxima verossimilhança (ML), utilizando-se o programa PhyML v. 3.0 (Guindon & Gascuel, 2003); inferência bayesiana (BI), com o programa MrBayes v. 3.1.2 (Huelsenbeck & Ronquist, 2001); e distância, através do algoritmo Neighbor-Joining (NJ) no programa MEGA 5. A escolha dos métodos adotados neste estudo foi baseada nas vantagens e desvantagens de cada um. Na construção de filogenias, os métodos discretos (ML e BI) analisam os sítios separadamente, já os métodos de distância utilizam uma matriz de distância estimada para cada par de sequências, ou seja, reduzem a informação a um único valor.

Como os ERVs analisados neste estudo estão presentes em diferentes hospedeiros, deve-se considerar que diferentes processos evolutivos podem estar agindo ao longo dos sítios. Então, se reduzirmos a variabilidade das sequências a valores únicos, estaremos perdendo a informação sobre processos evolutivos contida em cada sítio. Por serem rápidos e simples, os métodos de distância foram utilizados como triagem para verificar

a qualidade do alinhamento e a presença de sequências discrepantes. As inferências filogenéticas abordadas neste estudo estão essencialmente baseadas em ML e BI.

Para reconstruções filogenéticas por ML no PhyML, a filogenia foi computada a partir de uma árvore *neighbor-joining* (NJ) e a busca da melhor árvore foi realizada através do método heurístico NNI (*Nearest Neighbor Interchange*). O suporte dos ramos foi avaliado pelo método de *bootstrap* com 200 replicações. Os modelos de substituição utilizados nas construções filogenéticas foram obtidos com o programa Jmodeltest v.0.1.1 (Posada, 2008) para os nucleotídeos e com o programa ProtTest software v. 2.4 (Abascal *et al.*, 2005) para as proteínas, ambos de acordo com o critério Akaike (Sullivan & Joyce, 2005), ou Akaike com correção (AICc) nos casos de tamanho amostral pequeno. Optou-se pelo critério de informação Akaike porque é uma das técnicas estatísticas que seleciona um modelo mais complexo somente quando necessário, reduzindo o erro nas estimativas.

Para análise bayesiana no Mr Bayes, foram utilizadas quatro cadeias de Markov (uma fria e três aquecidas) percorridas por  $2 \times 10^6$  gerações, amostrando árvores a cada 100 gerações e, posteriormente, descartando-se 25% das amostras iniciais (*burn-in*). A probabilidade posterior para os clados foi reportada em uma árvore consenso (*majority rule consensus tree*). Os modelos evolutivos utilizados nas análises de proteína foram sugeridos pelo programa.

Em todas as árvores a raiz foi colocada no ponto médio, exceto nas filogenias das Figuras 2, 11 e 15 que apresentam árvores não enraizadas.

### **Networks (redes filogenéticas)**

A história de um gene nem sempre será melhor representada por uma árvore (Holmes *et al.* 1999) onde qualquer nó interno conecta tão somente três ramos (bifurcação) e todas as partições (*splits*) serão compatíveis, significando que todas as partições possíveis podem ser representadas em uma única árvore. Diferentemente das árvores filogenéticas, as *networks* podem conter um conjunto de *splits* incompatíveis e são uma excelente ferramenta para representar alguns tipos de relacionamentos evolutivos, tais como politomias (por exemplo, uma evolução rápida de HIV em pacientes), recombinação, hibridização e transferência de genes, porque não forçam os dados como na construção uma árvore (Holmes *et al.*, 1999; Moulton & Huber, 2009).

Além disso, as *networks* são muito úteis porque permitem explorar e entender melhor os dados, possibilitando visualizar melhor a sua estrutura e identificar padrões não representados em uma árvore filogenética.

Neste estudo, construímos *networks* com o programa Splitstree4 v.4.12.3 (Huson & Bryant, 2006), usando o método *split decomposition* e tomando por base a distância genética corrigida, conforme o modelo obtido no programa Jmodeltest ou modelos descritos na literatura (Holmes *et al.* 1999).

#### **4.3.6. Seleção**

A adaptação por seleção natural é um dos mais importantes processos na biologia, podendo explicar, em parte, a complexidade e diversidade biológica. Para detectar a seleção, utilizam-se geralmente diferentes tipos de informações que são influenciadas pelo tipo de sequências representadas no alinhamento (Pybus & Shapiro, 2009). Por exemplo, as diferenças observadas em populações são resultantes do atual polimorfismo (recente em termos evolutivos), já na comparação entre espécies, as diferenças foram fixadas em eventos que ocorreram no passado mais longínquo (Kyazhimskiy & Plotkin, 2008). Para analisar os diferentes tipos de informações, aplicam-se diferentes métodos.

Há métodos estatísticos que investigam a seleção através de um alinhamento de sequências com base na frequência relativa de sítios polimórficos (Pybus & Shapiro, 2009). Essa estatística é comparada com valores esperados para um modelo de evolução neutro. Nesses métodos assumem-se as seguintes premissas: tamanho efetivo da população e taxa de mutação constante; ausência de recombinação e migração; e acasalamento não direcionado (Page & Holmes, 1998). A vantagem desses métodos é a sua simplicidade; contudo, a rejeição da hipótese nula não significa necessariamente ação de seleção, pois é possível que haja violação de algum dos pressupostos mencionados. O método mais conhecido é o D de Tajima (Tajima's D) que é baseado na média de sítios que diferem entre cada par de sequências ( $\pi$ ) e o número de sítios segregantes (S), isto é, não fixados dentro de uma amostra de sequências (Tajima, 1989). No modelo neutro de evolução  $\pi$  e S são gerados sob os mesmos processos, resultando em um D de Tajima próximo a zero. Valores negativos ocorrem quando há seleção positiva, processo em que as sequências descendem de um ancestral comum recente, refletindo um excesso de sítios polimórficos ( $S > \pi$ ); ao passo que valores

positivos podem significar seleção estabilizadora. Neste estudo, o D de Tajima foi calculado com o programa DNAsp 5.1 (Librado & Rozas, 2009) em situações nas quais se buscou examinar se as sequências similares, pertencentes à mesma espécie, estavam evoluindo de modo neutro.

Os métodos dN/dS ( $\omega$ ) são baseados em mutações sinônimas e não sinônimas de uma sequência codificante. Devido à redundância do código genético, mutações silenciosas não alteram o aminoácido. O padrão de substituição das mutações sinônimas e não sinônimas deve refletir uma ação de seleção natural. Como regra geral, desvio de dN/dS=1 significa seleção; dN/dS < 1 significa seleção negativa, ou seja, substituições não sinônimas acumulando-se mais lentamente que as sinônimas; e dN/dS>1 seleção positiva. Como nem toda a seleção sinônima pode ser considerada neutra, esses métodos podem falhar em capturar alguns tipos de seleção, como, por exemplo, seleção preferencial por determinados códons (*codon usage*). Esses métodos são mais empregados para analisar sequências de diferentes espécies onde os polimorfismos intraespecíficos são ignorados (Kyazhimskiy & Plotkin 2008).

Com objetivo de determinar se clados de ERVs *Gamma-like* estavam sobre pressão seletiva foi selecionado um fragmento de 456 pb, que inclui 1°- 4° domínio da RT. Esse fragmento foi escolhido por apresentar menor número de inserções e deleções responsáveis por mudanças de fase de leitura e o menor número de códons de terminação. Os alinhamentos foram editados manualmente no programa MEGA5, colocadas em fase de leitura (considerando a hipótese de homologia das regiões codificadoras) e, em seguida, os códons de terminação foram removidos, pois esse evento não é admitido em uma sequência de proteína funcional. Para cada conjunto de dados foi construída uma filogenia, pelo método de máxima verossimilhança no programa PhyML, com base no modelo selecionado no programa Jmodeltest. Estas filogenias e os respectivos modelos foram utilizados para estimativas de seleção. Para o cálculo do coeficiente de seleção, utilizou-se o programa Hyphy (Pond *et al.*, 2005), no qual as estimativas de dN/dS são calculadas por meio da verossimilhança e de modelos explícitos de substituição de códons dentro de um contexto filogenético.

Antes de iniciar as análises de seleção foi verificada a presença de recombinação em posições específicas do alinhamento (*Breakpoints*) no programa Hyphy através de uma estimativa baseada na máxima verossimilhança. A seguir, foi estimada a taxa

global de seleção (que não varia de sítio para sítio ou entre ramos da árvore), permitindo uma aproximação da seleção da região analisada. A taxa global foi escolhida devido ao tamanho pequeno do fragmento analisado, o qual gerou filogenias com ramos internos de baixa resolução. A estimativa dessa taxa é mais simples, contendo menor número de parâmetros.

#### **4.3.7. Recombinação**

A recombinação é um dos mais importantes mecanismos geradores de diversidade. Encontrar um sinal de recombinação homóloga muitas vezes é um desafio, pois a grande maioria dos eventos recombinantes não deixa traços na molécula em que são gerados.

A observação de eventos discordantes na filogenia é o primeiro passo para evidenciar eventos de recombinação. Nesse caso o uso das *networks*, como comentado anteriormente na Seção 4.3.5 sobre reconstrução filogenética, provém uma excelente ferramenta na identificação do padrão de evolução reticulada. O método *split decomposition*, implementado no programa SplitsTree4, é muito eficiente na detecção de sinais filogenéticos conflitantes (Holmes *et al.* 1999). No método *split decomposition* as partições (*splits*) são inferidas da matriz de distância e, dependendo dos valores obtidos, pode ser produzido um diagrama semelhante a uma árvore filogenética ou a uma árvore reticulada. Entretanto, o padrão de evolução em rede não pode ser creditado exclusivamente a eventos recombinantes, pois ele também pode refletir polimorfismos causados por mutações recorrentes. Existem vários algoritmos genéticos que testam recombinação, o que reflete de certa forma, a dificuldade de avaliá-la (Lemey & Posada, 2009). O índice de homoplasia calculado no teste *Phi* (Bruen *et al.*, 2006), que mede o grau de similaridade entre sítios proximamente ligados, é um poderoso método para detectar recombinação. Ele se mostra eficiente tanto em sequências bastante relacionadas quanto em sequências distantes, além de apresentar a facilidade de estar implementado no pacote do programa SplitsTree4.

## 5. RESULTADOS

### 5.1. Análise dos clones

Foram sequenciados e analisados 236 clones, sendo 85 clones similares a retrovírus (Tabela 1). Todas as espécies de felídeos estudadas possuem sequências de ERVs. A origem retroviral desses clones se sustenta pela presença de sequências características da região da transcriptase reversa/protease e valor de  $E < 10^{-10}$  (em relação a ERVs depositados no GenBank). A maioria das sequências de ERVs ( $n = 82$ ) apresentou códons de parada (*stop codons*), mutações que ocasionam a mudança da fase de leitura ou grandes deleções, indicando que esses retrovírus são defectivos e, portanto, não funcionais (Apêndice B).

**Tabela 1.** Número de clones similares a ERVs encontrados nas espécies *Panthera onca* (OC), *Puma concolor* (Pco), *Puma yagouaroni* (Py), *Leopardus colocolo* (Lco), *Leopardus geoffroyi* (Lg), *Leopardus pardalis* (Lp), *Leopardus wiedii* (Lwi), *Leopardus tigrinus* (Lti).

Retroelemento	OC	Pco	Py	Lco	Lg	Lp	Lwi	Lti
<i>Gamma-like</i>	5(4)	16(7)	10(5)	10(3)	14(9)	14(6)	6(6)	5(4)
Classe I (NI)	0	0	0	0	0	0	1	0
Classe II (NI)	0	0	3(1)	0	0	0	0	0
RV-Cougar	0	0	0	0	1	0	0	0
Total ERVs	5	16	13	10	15	14	7	5
Total clones	25	31	25	25	33	32	40	25
%	20	51,6	52	40	45,4	43,7	17,5	20

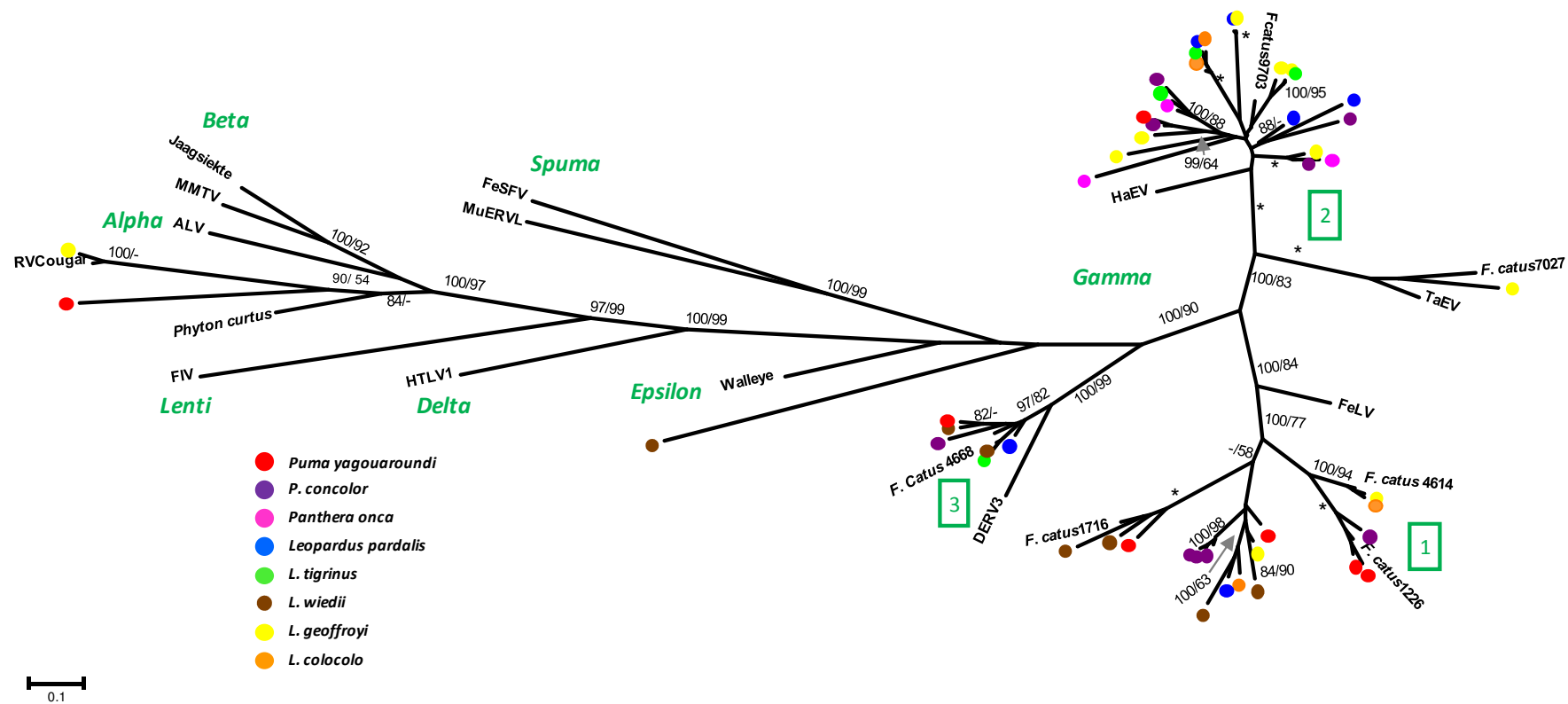
NI, não identificado; números entre parênteses representam o número de linhagens (definidas com base em uma distância genética observada intragrupo  $\leq 5\%$ ).

Os resultados do BLASTn e BLASTx e das análises filogenéticas demonstram que a maioria dos clones está estreitamente relacionada com o gênero *Gammaretrovirus* (*Gamma-like*) (Tabela 1 e Figura 2).

A proporção entre número de clones com sequências de ERVs em relação ao número total de clones sequenciados foi maior nas espécies *Puma concolor* e *Puma yagouaroundi* (pertencentes à linhagem Puma), apresentando valores percentuais muito próximos (Tabela 1). As espécies da linhagem Jaguatirica, por sua vez, demonstraram grande variação com relação ao número de clones que continham sequências de ERVs (17,5-43,7%). Na espécie *Panthera onca* (representante da linhagem Pantera), foram obtidos apenas 20% de clones positivos (Tabela 1).

Quando as sequências com distância genética menor que 5%, ou seja, com similaridade de 95% ou mais são agrupadas em linhagens (Tabela 1), percebe-se que o número mais alto de ERVs encontrados em algumas espécies estava relacionado à amostragem de clones similares. Os resultados demonstrados na Tabela 1 sugerem que as espécies de felídeos silvestres brasileiros apresentam, em média, 5,5 linhagens de retroelementos similares a retrovírus inseridas nos seus genomas.

A filogenia com parte da região da RT (Figura 2) sugere que os ERVs *Gamma-like* são divididos em três grandes grupos, resolução que recebe alto suporte estatístico tanto na análise Bayesiana como na ML. Dentro desses grupos, os ERVs pertencentes às diferentes espécies apresentam-se distribuídos de forma heterogênea, o que dificulta a observação de qualquer padrão coevolutivo. É possível observar também algumas sequências isoladas que se agrupam próximo a *Beta* e a *Alpharetrovirus* e uma sequência próxima a *Epsilonretrovirus*. Isso indicaria que ancestrais desses felídeos também eram infectados por outros gêneros além do *Gammaretrovirus*.



**Figura 2.** Filogenia Bayesiana demonstrando a diversidade dos clones amostrados nas espécies de felídeos silvestres brasileiros. A análise foi baseada em 161 aminoácidos da região da transcriptase reversa e no modelo JTT+G. Os valores próximos aos ramos, à esquerda, representam a probabilidade posterior (> 0.8) da filogenia BI, e à direita, valores de *bootstrap* (>50%) correspondente a filogenia ML. O asterisco representa o valor de 100% em ambas as análises. A lista dos clones encontra-se no Apêndice B e a dos demais ERVs, no Apêndice A. *Felis catus* está identificada pelos quatro últimos algarismos das sequências listadas no Apêndice E. Os números 1-3 dentro de caixas representam três distintos clados de ERVs *Gamma*-like.



## 5.2. SINEs identificados nos fragmentos sequenciados

Os Clones LgEM10003 e PcoJO10006 possuem inserções de sequências similares a SINEs (Short Interspersed Nuclear Elements) identificadas através dos programas BLAST e BLAT; este último em conjunção com o programa RepeatMasker (Tabela 2). As duas inserções ocorreram no terceiro domínio da RT.

Além da similaridade revelada no BLAST, esses fragmentos apresentam tamanho compatível com um elemento SINE (236 pb e 232 pb, respectivamente), que em geral é menor que 500 pb (Symer & Boeke, 2010). As duas regiões promotoras da RNA polimerase III das sequências SINEs inseridas nos clones LgEM10003 e PcoJO10006, bem como as suas semelhanças com a sequência de um SINE proveniente do gato doméstico (SINEC\_Fc2), são demonstradas na Figura 3.

A busca por sequências similares no NCBI, através do programa BLASTn identificou dois clones com sequências similares ao retroelemento LgEM10003: o retrovírus endógeno RV-cougar, originado de *Puma concolor*, e RV-domestic cat do gato doméstico. Ambos são provenientes do estudo de Grifford (2005). A divergência genética (distância p) entre LgEM10003 e RV-cougar ( $d=0.023$   $SE=0.005$ ) e RV-Domestic cat ( $d=0.11$   $SE=0.012$ ) indica alta similaridade entre essas sequências. Entretanto, as sequências de RV-cougar e RV-Domestic cat não apresentam a inserção do SINE.

O clone LgEM10003 pode ser classificado como pertence à classe II de retrovírus e, aparentemente, é um elemento bastante antigo, visto que a sequência apresenta muitos códons de parada e *indels* que provocaram mudanças na fase de leitura (Apêndice B). Da mesma forma, nenhuma sequência de SINE foi encontrada em sequências similares a PcoJO10006 provenientes deste estudo. O clone PcoJO10006 pertence à classe I de retrovírus (como será descrito na Seção 5.3.2 e na Figura 6), apresenta apenas um códon de parada e podendo, portanto, ser considerado como um elemento recente.

**Tabela 2.** Fragmentos heterólogos similares a SINEs inseridos nos clones PcoJO10006 e LgEM10003 (Segmento 2) e demais similaridades encontradas nos segmentos 1 e 3 das sequências.

Clone	Segmento 1	Segmento 2	Segmento 3
<b>PcoJO10006</b>	<b>1-167pb</b> Gibbon ape leukemia virus ( U60065.1) I=87% C=28% e=2e-06 Blastn	<b>168-399pb</b> SINEC_Fc2 [chrX:61882383-61882587]* I=83.7% Blat/ RepeatMasker	<b>400-1211pb</b> Murine leukemia vírus (X99925.1) I=73% C=98% e= 6e-131 Blastn
	<b>LgEM10003</b>	<b>1-583pb</b> RV-Cougar (AY820105.1) I=97% C=99% e=0.0 Blastn	<b>584-819pb</b> <i>Otocolobus manul</i> (AY518631)-SINE I=96% C=90 % e= 5e-93 Blastn

I=identidade, C=cobertura, e= *e value*. Os número de acesso do GenBank aparecem entre parênteses, o asterisco representa a localização no genoma de *Felis catus* (montagem versão catChrV17e). Blastn e Blat/RepeatMasker foram os programas utilizados nas buscas.

```

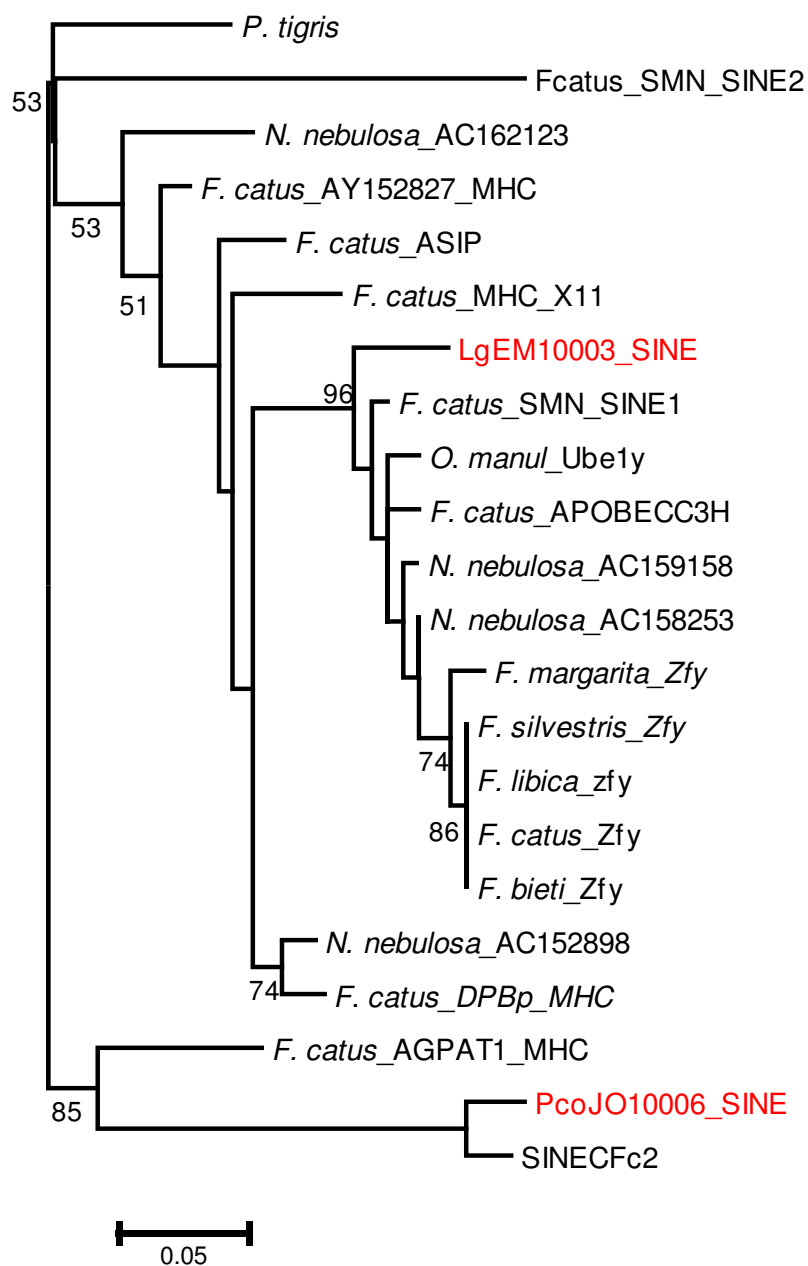
          *      20      *      40      *      60      *      80
LgEM10003 : GGGGCGCCTGGTGGCGCAGTCGGTTAAGCGTCCAACCTC-GGCCAGGTCATGATCTCGCAGTCCGTGAGTTCGAGCCCC : 79
PcoJO10006 : GGGGTGCCATGATGGCTCAGTCCATTAAGCAACTGACTTCAGGTCAGGTCATGATCTCACTGTTGGTGGGTTCAAGCCC- : 79
SINEC_Fc2  : GGGGTGCCATGATGGCTCAGTCCATTAAGCGACTGACTTCAGGTCAGGTCATGATCTCACTGTTGGTGGGTTCAAGCCC- : 79
          _____
                    A box
          *      100     *      120     *      140     *      160
LgEM10003 : GTGTCGGGCTCTGGGCTGATGGCTCAGAGCCTGGAGCCTGTTTCTGATTCGTGTCTCCCTCTCTCTGCCCCCTCCCCC : 159
PcoJO10006 : GTGTCGGGCTCTGTGCTGACAGCTATGGGCTTGGAGCCTGTTTCCGGCTCAGTATCTCCCTCCCT---GCTCCTCCCCC : 155
SINEC_Fc2  : GTGTCGGGCTCTGTGCTGACAGCTATGGGCTTGGAGCCTGTTTCCGGTTCGTATCTCCCTCTCT---GCTCCTCCCCC : 155

          *      180     *      200     *      220
LgEM10003 : GTTCATGCTCTGTCTCTCTCTG--TCCCAAAAATAAATAAACGTTGAAAAAAAAATTTTTTTT : 219
PcoJO10006 : ACTTGCACCTCTGTCTCTCTCTGCCTCTTGAAAACAAATAAATGTTAAAAAATTTTTTAAA : 217
SINEC_Fc2  : ACTTGCACCTCTGCCTCTCTCTGCTTCTTGAAAACAAATAAATGTTAAAAA----- : 205

```

**Figura 3.** Alinhamento demonstrando um fragmento das sequências dos clones LgEM10003 e PcoJO10006 . As linhas delimitam os Box A e B do promotor para RNA polimerase III (com base em Pecon-Slattery *et al.* 2000). A sequência SINEC\_Fc2 está descrita no Apêndice C.

A filogenia representada na Figura 4 corrobora os resultados descritos na Tabela 2 agrupando os SINEs inseridos nos retroelementos LgEM10003 e PcoJO10006 junto com sequências de SINEs provenientes de outras espécies de felídeos. As sequências desses dois novos elementos identificados agrupam-se em clados separados, indicando que as cópias integradas são derivadas de distintos progenitores, demonstrando dois eventos distintos de integração.



**Figura 4.** Árvore filogenética de máxima verossimilhança que demonstra o relacionamento evolutivo entre os dois elementos isolados neste estudo (em vermelho) com SINEs de outras espécies de felídeos, conforme se encontra no Apêndice C. A filogenia foi construída com base em um alinhamento de 189 pb (com 48 sítios informativos para parcimônia), após excluída a cauda poli-A/T. A árvore foi baseada no modelo GTR+G e enraizada no ponto médio. Números próximos aos ramos representam os valores de *bootstrap* (> 50%).

### 5.3. Análises filogenéticas

#### 5.3.1. Alinhamentos

Na maioria das análises, foi utilizada uma matriz de proteína composta pelo fragmento da região PR e RT semelhante ao exemplificado na Figura 5. O motivo da utilização de sequências de proteína nestas análises deve-se à evidência de saturação de substituições (ver Seção 4.3.3) encontrada na maioria dos alinhamentos de sequências de nucleotídeos (gráficos não mostrados). Uma região variável entre PR e RT de aproximadamente 100 aminoácidos (variável de acordo com o ERV) foi retirada dos alinhamentos devido à dificuldade de alinhar posições homólogas. Essa estratégia foi aplicada para os alinhamentos das filogenias representadas pelas Figuras 6 a 10, 13, 14, 16, 20 e 21. Entretanto, na comparação entre grupos menos divergentes, por exemplo, o grupo *Myotis* (Figura 17), manteve-se todo o fragmento da RT sequenciado. Aminoácidos não alinhados pelo programa DIALIGN2, códons de parada, bem como as inserções necessárias para manter a fase de leitura foram considerados como sendo caracteres desconhecidos. Como já comentado na metodologia, todos os alinhamentos foram refinados manualmente. Nos grupos menos divergentes foi igualmente possível analisar alinhamentos de nucleotídeos. Dependendo do conjunto de sequências analisadas, o total de sítios dos alinhamentos sofreu variações, pois as sequências obtidas neste estudo variaram bastante em tamanho (Apêndice B).

PR/RT

1

```

*      *      *      *      *
20    40    60    80    100   120
LcoJo10001 : XSQVPLGQKTATILGATGWV-MQWPFCA-QQCE--L-GRHKVVHFEFLYLPDCPI PLL-GQDLSKIPGAQPQRORQYLLPLEARIGIEEH-LTKLREAGILIECQSAWNI PLLPVKKPG : 115
PcoJo10006 : TSLVVPRLRQMANIIGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDRPI RLL-GRDLSKILGAQPQRORQYPLPLEARIGLQEN-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
LPJo10003 : TSQVAPLSQKTATILGATGTQA-MXQPFCA-QQCE--LXGGHKIRHEFLYLDPCLIP LL-GRDLSKILGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 116
OCJo10001 : TSLVAPLQKTATILGATGTQA-MQWPFCA-QQGA--L-GGHKARHEFLYLDRDCPI PLL-GRDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
OCJo10005 : TSLVVPLSQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI PLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
PcoJo10007 : TSLVVPSSQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI PLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
LPJo10002 : ISXVVPFSQKMATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI PLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
LgCt10005 : TSPVAPLRQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHK--HEF-YLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 112
Pcatus9703 : TSPVAPLSQKMATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI PLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
PYCt10001 : TSPVAPLSQKMATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI PLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
PcoJo10009 : TSPVAPLSQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI PLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
LgJo100010 : TSLVAPFSQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDX-SPLL-DRDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
LgJo10002 : XXXXXXXXQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDX-SPLL-DRDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
LtiJo10009 : TSLVAPFSQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDX-SPLL-DRDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
LPJo10004 : TSLVAPFSQKMATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI PLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
OCJo10002 : TSPVAPFSQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI PLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
LgJo10016 : TSLVAPFSQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI PLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
LtiJo12009 : TSOVAPFSQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHEVQHEFLYLPDCPI RLL-CXDVLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
PcoJo12001 : TSQVAPFSQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHEVQHEFLYLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 116
HaEV : TSPVAPFSQKMATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI PLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
LgCt10007 : TSPVAPLSQKTATILGATGTQA-MQWPFCA-QQCE--LXGGHTIRYEFYLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 109
MiEVII : TSLVVPFSQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
MeEVII : TSPVVPFSQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHEVQHEFLYLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
LcoJo10002 : TSOVVPFGQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI PLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
LtiJo10005 : TSOVVPFSQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI PLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
LPJo3 : TSOVVPFSQKTATILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI PLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
Pcatus7027 : TQKVPVPLSGREVITILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI RLL-GRDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
TaEV : TQKVPVPLSGREVITILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI RLL-GRDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
LgJo10008 : THKVPVPLSGREVITILGATGTQA-MQWPFCA-QQCE--L-GGHKVVHFEFLYLPDCPI RLL-GRDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
BoEV : TPKVPEPLFRKTTAIFGATG-MQ-TXKPFCS-PXQXVXG-GGHQVHFEFLYLPDCPI RLL-GRDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
LPCT10005 : XXXXXXXXXXXXXXXXXXXXINGTSPXIPCTLL-V-LTCC--L-DGPHFSHFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
LwiJo10004 : PYSYGPPTPSSSVVINGTSPXIPCTP-V-LTCC--L-DGPHFSHFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
Pcatus4668 : PYSYGPPTPSSSVVINGTSPXIPCTP-V-LTCC--L-DGPHFSHFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
LwiJo12004 : PYSYGPPTPSSSVVINGTSPXIPCTP-V-LTCC--L-DGPHFSHFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
LtiJo10006 : PYSYGPPTPSSSVVINGTSPXIPCTP-V-LTCC--L-DGPHFSHFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
PcoJo10004 : PYSYGPPTPSSSVVINGTSPXIPCTP-V-LTCC--L-DGPHFSHFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
PYJo10003 : PYSYGPPTPSSSVVINGTSPXIPCTP-V-LTCC--L-DGPHFSHFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
DERV3 : PYSYGPPTPSSSVVINGTSPXIPCTP-V-LTCC--L-DGPHFSHFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
ERV-Fc1 : PYSYGPPTPSSSVVINGTSPXIPCTP-V-LTCC--L-DGPHFSHFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
HERV-H : PYSYGPPTPSSSVVINGTSPXIPCTP-V-LTCC--L-DGPHFSHFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 114
Pcatus1716 : KQDGPPTRRGATLVQVGTGSGT-CRST--EXRQVD--L-GRETVTHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 115
LwiJo10002 : XXXXGPTRRGATLVQVGTGSGT-CRST--EXRQVD--L-GRETVTHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 90
LwiJo10003 : KQDGPPTRRGATLVQVGTGSGT-CRST--EXRQVD--L-GRETVTHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
PYJo10004 : XXXXGPTRRGATLVQVGTGSGT-CRST--EXRQVD--L-GRETVTHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
FERVmlu2 : RTPGKGLASKSSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
FERVmlu1 : RTPGKGLASKSSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
MeEVI : LEPMGKLSGEEVSVQGANSTER-HRRT--E-RTLT--L-ASGVVXKRFVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
PYJo10002 : LQPGKGLASKSSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
Pcatus1226 : LQPGKGLASKSSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
DERV1 : LQPGKGLASKSSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
VuEV : LQPGKGLASKSSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
PYJo10001 : LQPGKGLASKSSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
PcoJo12004 : LQPGKGLASKSSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
LgJo10006 : IRPHGKISVKSSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
LcoJo10001 : IRPHGKISVKSSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
Pcatus4614 : IRPHGKISVKSSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
PoEV : LQPGKGLASKSSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
KoRo : XXXXGPTRRGATLVQVGTGSGT-CRST--EXRQVD--L-GRETVTHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
RD114 : TKANGPLSSRTSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
BaEV : TKANGPLSSRTSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
FeLV : TRPDGPLSSRTSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
MDEV : TRPDGPLSSRTSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113
MiEVI : NQKDGMSKSSWVQATGMSQ-YSWTT--R-RTVD--L-GTGRVSHSFLVLPDCPI RLL-GQDLSKIPGAQPQRORQYPLPLEARMGIQEH-LTKLREAGILIECQSAWNT PLLPVKKPG : 113

```

Figura 5. Alinhamento das seqüências de retroelementos similares a *Gammaretrovirus* utilizadas na construção da filogenia da Figura 20. Os números 1-5 representam os domínios da RT com base em Xiong & Eickbush (1990).

(Continua)



### 5.3.2. Análises intraespecíficas

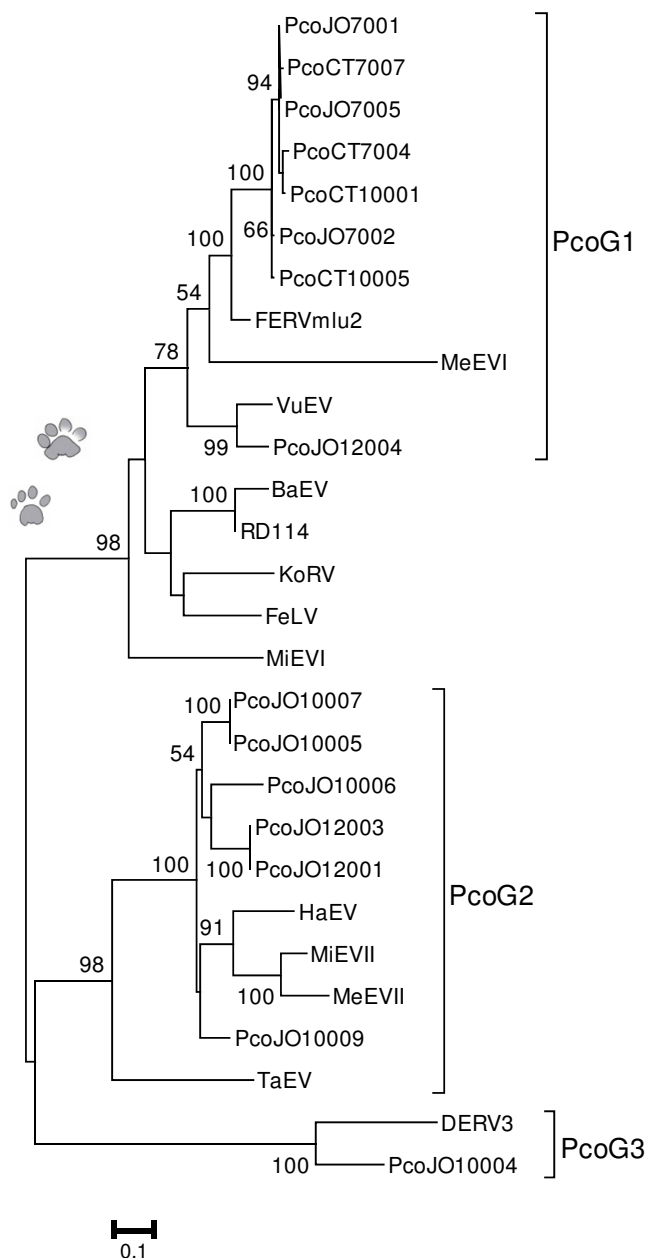
A seguir, resume-se a história evolutiva presumida dos retroelementos recuperados em cada espécie estudada. Os ERVs *Gamma-like* foram nomeados G1, G2 e G3, a fim de acompanhar o padrão geral descrito na Figura 2.

#### *Puma concolor*

No *Puma concolor*, foram encontrados 16 retroelementos, todos *Gamma-like* (Tabela 1). Os relacionamentos evolutivos entre esses clones foram recuperados através de uma filogenia construída com um alinhamento de 237 aminoácidos correspondente a um fragmento das regiões PR e RT. O clone PcoCT10002 foi excluído do alinhamento por apresentar somente a região da protease.

A Figura 6 mostra que os 15 ERVs *Gamma-like* analisados são divididos em três grupos denominados PcoG1 a PcoG3. O grupo PcoG1 está dividido em dois ramos: um deles contém sete sequências muito similares a FERVmlu2, um retrovírus endógeno proveniente do gato doméstico similar ao do morcego *Myotis lucifugus* (ver a Seção 5.3.3); o outro é representado por uma sequência que se agrupou próxima ao ERV relacionado com o vírus da leucemia murina (MLV), oriundo de *Vulpes vulpes* (VuEV). O Grupo PcoG2 é composto por quatro ramos proximamente relacionados, com sequências pertencentes a um subgrupo de MLVs, que foram descritos por Tristem (Tristem *et al.*, 1996). O grupo PcoG3 é representado por um ramo da árvore em que foi capturada apenas uma sequência. Essa sequência apresenta 67% de identidade (BLASTn) com DERV3, um ERV proveniente do cão. Todas as sequências obtidas de *P. concolor* são defectivas, pois apresentam códons de parada e/ou *indels* que provocaram mudança na fase de leitura ou grandes deleções da PR e de parte da RT.

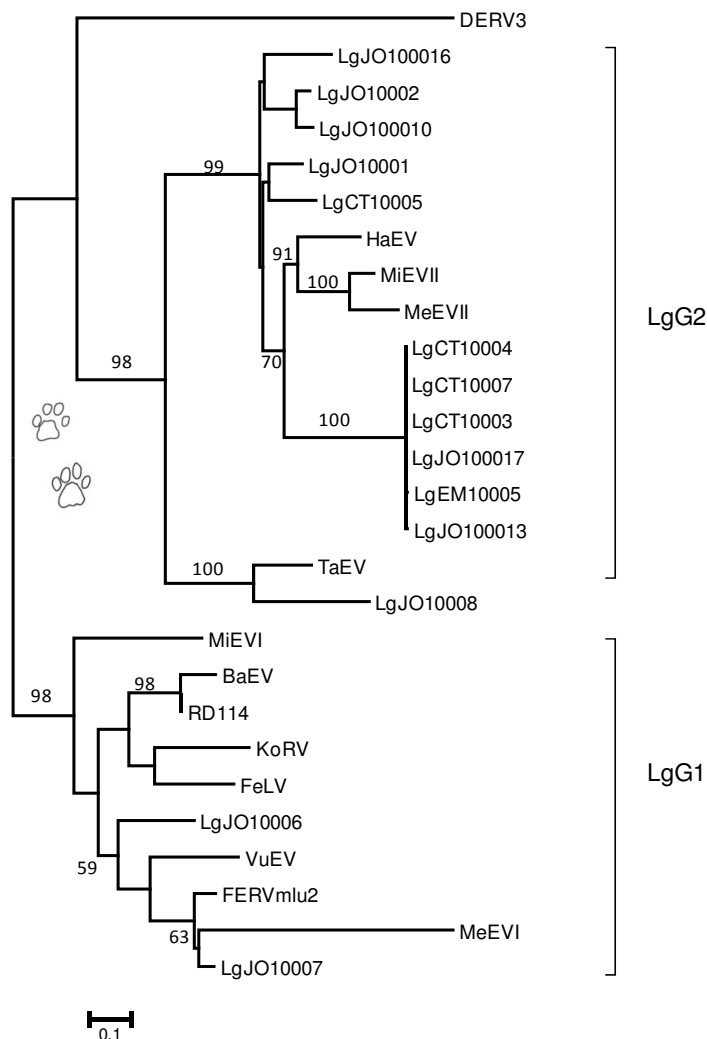




**Figura 6.** Relacionamentos filogenéticos entre retroelementos *Gamma-like* de *Puma concolor* com seqüências conhecidas de ERVs (Apêndice A). A Filogenia (ML) foi baseada em 237 aminoácidos segundo modelo JTT+G+F. Os números próximos aos ramos representam os valores (%) de *bootstrap* > 50%.

### *Leopardus geoffroyi*

Dos 14 clones similares a ERVs encontrados na espécie *L. geoffroyi*, 13 são *Gamma-like*, conforme demonstrado na Tabela 1 e na Figura 7.



**Figura 7.** Relacionamentos filogenéticos entre retroelementos similares a *Gammaretrovirus* encontrados em *L. geoffroyi* e seqüências conhecidas de ERVs (Apêndice A). A Filogenia (ML) foi baseada em 242 aminoácidos, segundo modelo JTT+G+F. Os números próximos aos ramos representam os valores de *bootstrap* > 50%.

No caso dessa espécie foram recuperados apenas dois subgrupos de seqüências (LgG1 e LgG2). O clado LgG1 é composto por duas seqüências (LgJO10006 e LgJO10007) que não apresentam códon de parada. Entretanto, a seqüência LgJO10007, que se agrupa com FERVmlu2, apresenta uma grande deleção da PR e de parte da RT; essa seqüência será apresentada novamente na Seção 5.3.3.

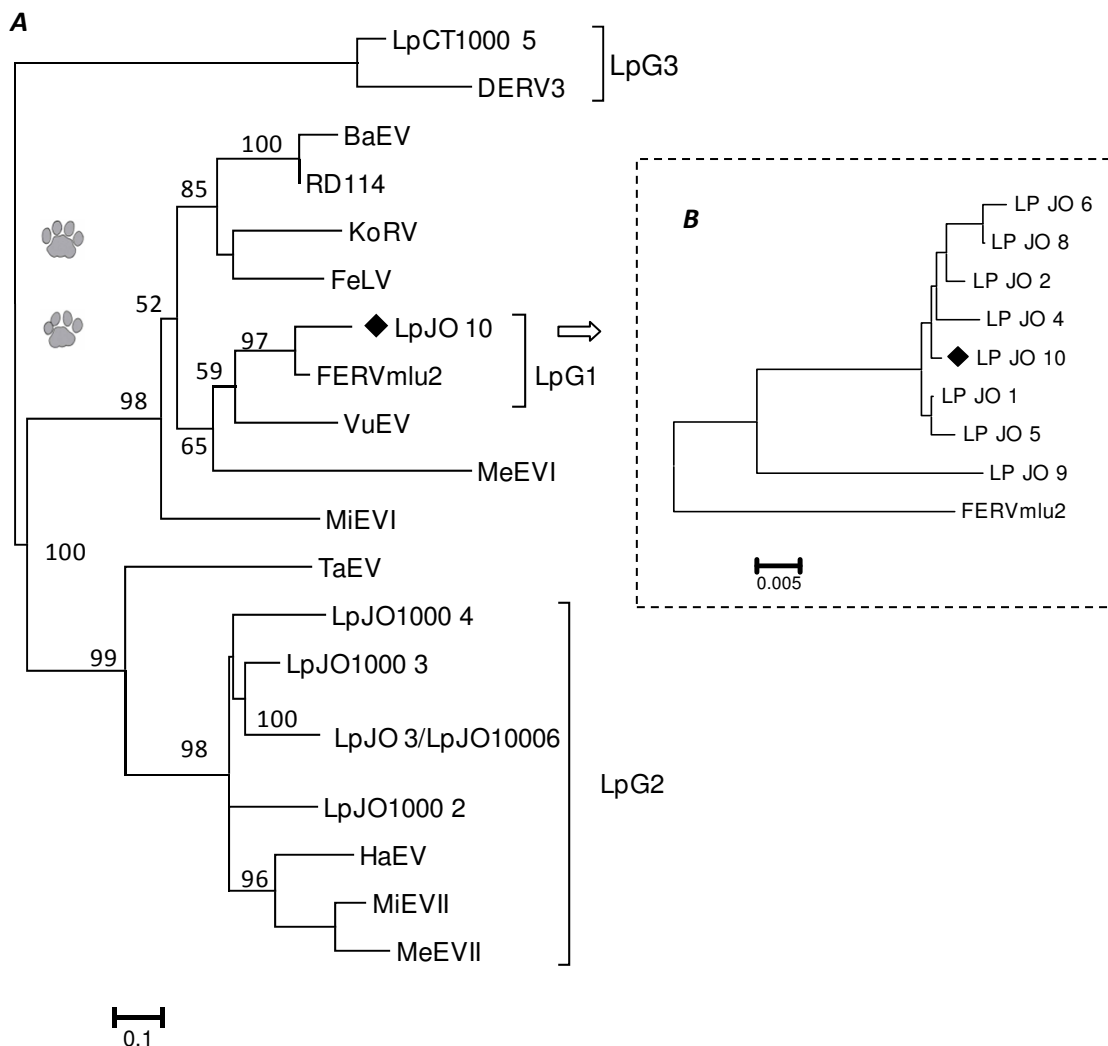
O LgG2 é formado por doze seqüências, todas defectivas, apresentando códon de parada. O clado que alberga a seqüência LgCT10004 é composto por seis clones muito similares ( $d = 0,003$ ;  $SE = 0,001$ ) que apresentaram, na região da RT, dois eventos de

deleção comuns a todas as sequências: um de sete aminoácidos e outro de trinta aminoácidos. As sequências de nucleotídeos dos clones LgCT10003 e LgCT0007 são redundantes, indicando baixa taxa de erro da PCR durante os processos de amplificação. O teste de neutralidade (Tajima's  $D = -0,504$ ;  $p > 0,1$ ), não significativo, sugere que esse grupo de sequências está evoluindo segundo um modelo neutro de evolução. A não violação do modelo de neutralidade também pode ser um indicativo que a população de sequências desse ERV está evoluindo de modo estável, sem expansão. Contudo, esse resultado deve ser tratado com cautela, devido ao pequeno tamanho da amostra.

O clone LgEM10003, como descrito anteriormente, apresenta 97% de similaridade de nucleotídeos com RV-cougar (GenBank AY820105) e 83% com RV-domestic cat (GenBank AY820046), os quais pertencem a retrovírus da classe II (Gifford, 2005).

### ***Leopardus pardalis***

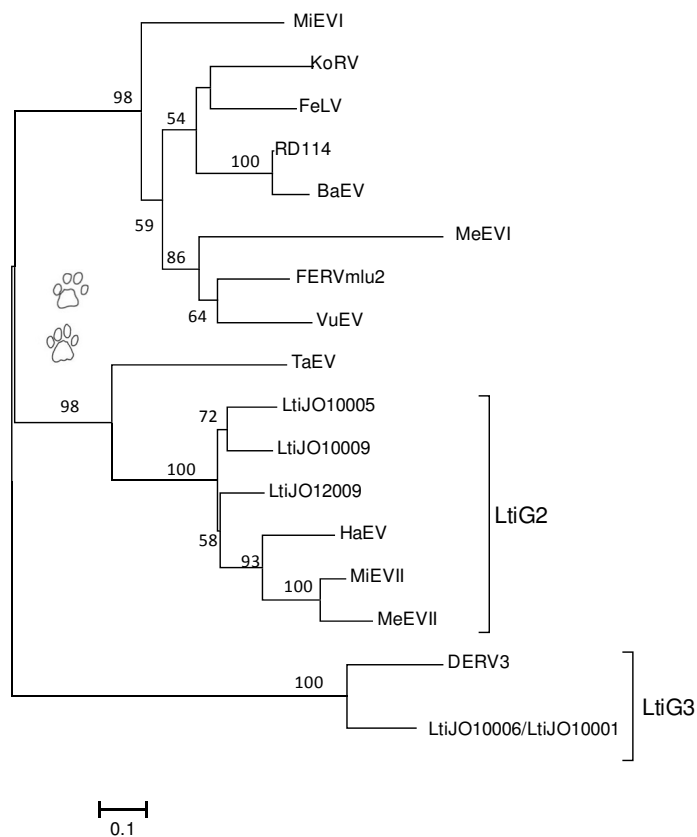
Os 14 clones obtidos em *L. pardalis* são *Gamma-like* (Tabela 1). Esses clones estão divididos em três subgrupos (LpG1 a LpG3) (Figura 8). O clado LpG1 está representado por apenas um grupo de sequências muito similares, proximamente relacionadas com FERVmlu2. Esse clado será apresentado com mais detalhes na Seção 5.3.3. O clado LpG2 apresentou cinco sequências, todas defectivas. A variação encontrada dentro desse clado, representada pela distância genética (distância  $p = 0,146$ ;  $SE = 0,012$ ), é maior que a apresentada no clado LpG1 (consulte a Figura 17B na Seção 5.3.3). Já no clado LpG3, foi recuperada apenas uma sequência, que se apresenta defectiva, com quatro códons de parada.



**Figura 8.** Relacionamentos filogenéticos entre retroelementos *Gamma-like* de *Leopardus pardalis* e sequências conhecidas de ERVs (Apêndice A). A. Filogenia (ML) foi baseada em 236 aminoácidos segundo o modelo JTT+G+F. Os números próximos aos ramos representam os valores de *bootstrap* > 50%. B. Árvore NJ foi inferida de um alinhamento de nucleotídeo com base na distância p. LpJO3 e LpJO10006 representam clones com sequências de aminoácido idênticas.

### *Leopardus tigrinus*

Os cinco clones obtidos na espécie *L. tigrinus* encontram-se divididos em dois cladogramas (LtiG2 e LtiG3) (Figura 9). Nenhuma sequência foi recuperada no grupo do FeLV (G1). As três sequências do clado LtiG2 são defectivas apresentando códons de parada ou inserções/deleções. A sequência LtiJO10006 do clado LtiG3 não apresenta códons de parada e *indels* que poderiam ocasionar mudança da fase de leitura.

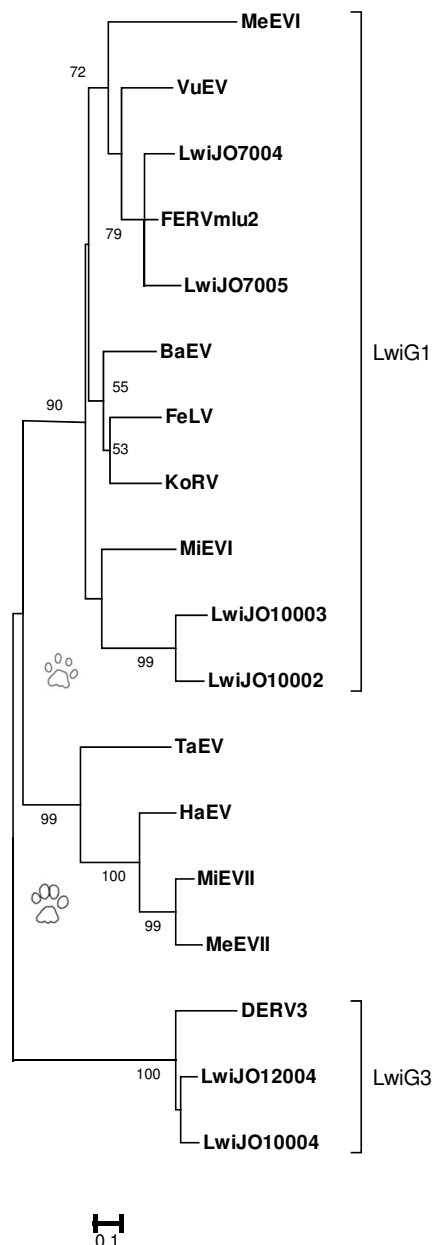


**Figura 9.** A filogenia (ML) foi baseada em 236 aminoácidos segundo o modelo JTT+G+F, demonstrando o relacionamento entre retroelementos de *Leopardus trigrinus*. As sequências de ERVs são descritas no Apêndice A. Os números próximos aos ramos representam os valores de bootstrap > 50%. LtiJO10006 e LtiJO10001 representam clones com sequências de aminoácidos idênticas.

### *Leopardus wiedii*

*Leopardus wiedii* foi a espécie com maior número de clones sequenciados (n = 40), entretanto, apenas sete (representando 17,5% dos clones sequenciados) apresentaram similaridade a ERVs (Tabela 1). Seis clones são *Gamma-like* e foram subdivididos em dois grupos (LwiG1 e LwiG3), conforme demonstrado na Figura 10. Nenhum representante do clado de *Tadarida* (TaEV), pertencente ao grupo G2, foi recuperado (ver Figura 2). LwiG1 é composto por dois clados: um se agrupou a FERVmlu2, e outro está proxicamente relacionado com MiEVI (um ERV do grupo dos MLVs provenientes da ovelha doméstica). Quando comparado com as espécies *L. pardalis*, *L. colocolo*, *P. concolor* e *P. yagouaroundi* (Figura 17B), o primeiro clado demonstrou a maior divergência genética (distância  $p = 0,089$ ; SE = 0,015) encontrada no grupo relacionado

com FERVmlu2. Entretanto, nesse clado foram recuperadas apenas duas sequências que são apresentadas novamente na Seção 5.3.3. O grupo LwiG3 está representado por duas sequências, sendo que LwiJO12004 não apresenta mutações ou *indel* na ORF parcial PR/RT.



**Figura 10.** Filogenia ML baseada em 239 aminoácidos segundo modelo WAG+G+F, demonstrando o relacionamento entre retroelementos de *Leopardus wiedii*. As sequências de ERVs são descritas no Apêndice A. Os números próximos aos ramos representam os valores de *bootstrap* >50%.

Nas buscas através do programa BLASTn no banco de dados nr do NCBI, o clone LwiJO7007 demonstrou 33% de identidade com *Walleye epidermal hyperplasia virus 2* (AAC59311.1), do gênero *Epsilonretrovirus*. O Programa Artemis identificou duas ORFs parciais, separadas por um códon de parada que, através de comparações com o BLASTp, evidenciou similaridades com o domínio da transcriptase reversa.

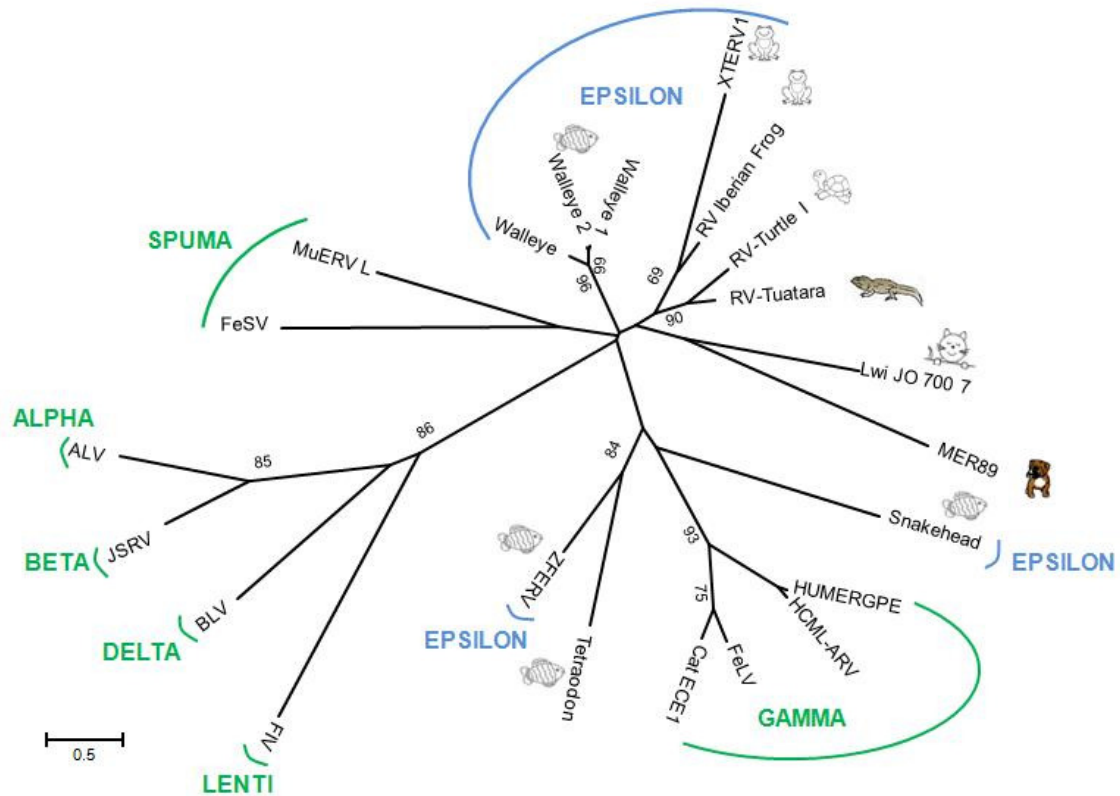
Utilizando-se o fragmento correspondente a essas duas ORFs como *query*, o programa RepeatMasker encontrou uma identidade de 69,5% com retroelementos classificados como ERV classe I. Com base nesses resultados, foi realizada uma nova busca com o programa BLASTn, comparando-se a sequência LwiJO7007 com o banco de dados nr do NCBI e restringindo-se a busca à espécie *Canis lupus familiaris* (taxid.: 9615). Um retroelemento localizado no cromossomo 11 (MER89) apresentou 72% de similaridade ( $e = 7e-37$ ), cobrindo aproximadamente a mesma região (posições 331 a 635) representada pelas ORFs parciais indicadas pelo programa Artemis.

Para compreender melhor os relacionamentos desse clone dentro da família *Retroviridae*, uma filogenia foi construída com as duas ORFs parciais (total de 106 aminoácidos da RT). Nela foram incluídos representantes de todos os gêneros de XRVs descritos, alguns ERVs e a sequência MER89 (Figura 11). O clone LwiJO7007 agrupou-se com a sequência de MER89. Esse clado está proximamente relacionado com o clado composto RV-Tuatara e RV-Turtle-I, ambos sem classificação e provenientes de répteis, e com RV-Iberian Frog e XTERV1, originários de anfíbios. XTERV1 foi descrito recentemente como *Epsilonretrovirus* (Sinzelle, *et al.* 2011). Entretanto, o suporte das topologias é muito baixo e, além disso, os relacionamentos no gênero *Epsilonretrovirus* não foram recuperados com clareza.

Com o propósito de verificar se sequências similares a LwiJO7007 estão presentes no genoma de outros mamíferos, foram realizadas buscas através do programa BLASTn, comparando-se essa sequência com o banco de dados WGS do GenBank e restringindo-se a busca a mamíferos (taxid.: 40674). Foram encontradas sequências muito similares à do clone LwiJO7007 em várias espécies de mamíferos (Figura 12). As identidades variaram de 96%, no gato doméstico, a 70%, no morcego *Pteropus vampyrus*.

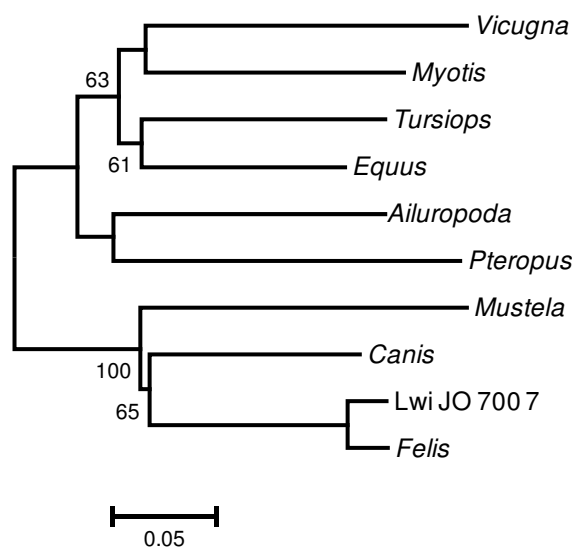
Portanto, sequências muito similares a LwiJO7007, aparentemente, encontram-se difundidas nos mamíferos placentários. Observando o tamanho dos ramos na filogenia

da Figura 12, as sequências parecem evoluir de modo homogêneo. A variação dentro dos dois cladogramas principais apresentados nessa figura é muito similar, confirmando o exposto acima. As sequências do clado *Mustela*, *Canis*, *Leopardus* (LwiJO7007) e *Felis* apresentaram distância p média de 0,18 (SE = 0,01) e as do clado *Vicugna*, *Myotis*, *Tursiops*, *Equus*, *Ailuropoda* e *Pteropus*, de 0,22 (SE = 0,09).



**Figura 11.** Relacionamentos filogenéticos entre o clone LwiJO7007 e representantes dos sete gêneros conhecidos de XRVs e ERVs. Estas sequências e respectivos números de identificação estão relacionados no Apêndice A e no texto. As topologias apresentadas são resultados de uma análise de máxima verossimilhança, utilizando-se o modelo RtREV+G+I. O alinhamento de 126 aminoácidos, incluindo os gaps, corresponde à região do 2° ao 5° domínio da RT. Números junto aos ramos representam os valores de *bootstrap* > 50 %.

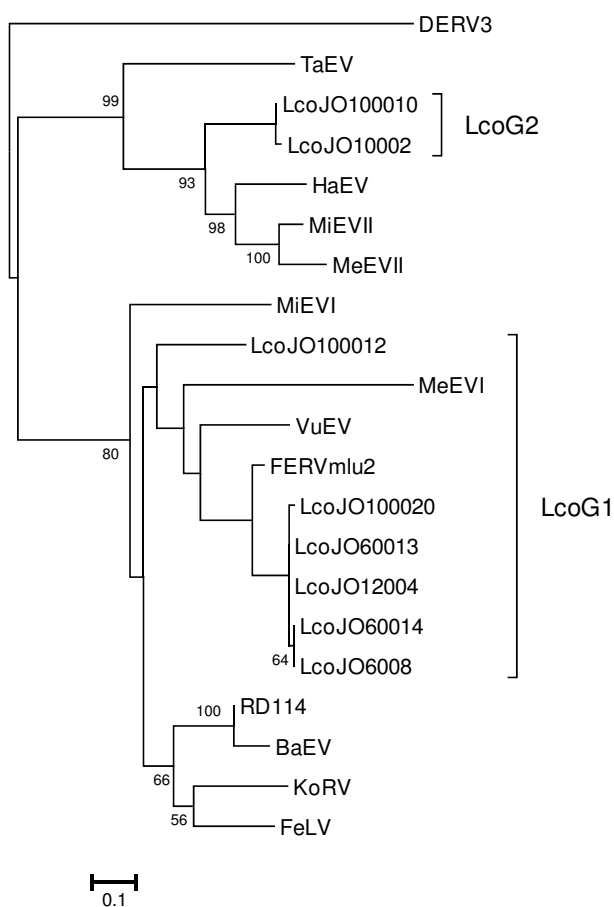




**Figura 12.** Árvore de máxima verossimilhança, baseada em um alinhamento de 658 nucleotídeos, segundo o modelo HKY+G, demonstrando o relacionamento entre as sequências similares ao clone Lwi JO 700 7 encontradas no genoma de alguns mamíferos placentários: *Felis catus* ACBE01155496.1, *Canis lupus familiaris* AAEX03002570.1, *Mustela putorius furo* AEYP01056564.1, *Pteropus vampyrus* ABRP01330907.1, *Ailuropoda melanoleuca* ACTA01106447.1, *Equus caballus* AAWR02025494.1, *Tursiops truncatus* ABRN01085982.1, *Myotis lucifugus* AAPE02021745.1, *Vicugna pacos* ABRR01374683.1

### ***Leopardus colocolo***

Na espécie *L. colocolo*, foram encontradas dez sequências de ERVs, todas defectivas (Apêndice B). Essas sequências são *Gamma-like* e apresentaram-se divididas em dois grupos: LcoG1 e LcoG2 (Figura 13). O Clado LcoG1 apresenta um grupo de cinco sequências que se agrupam com FERVmlu2 e são apresentadas mais detalhadamente na Seção 5.3.3. Não foram recuperadas sequências pertencentes ao grupo G3.



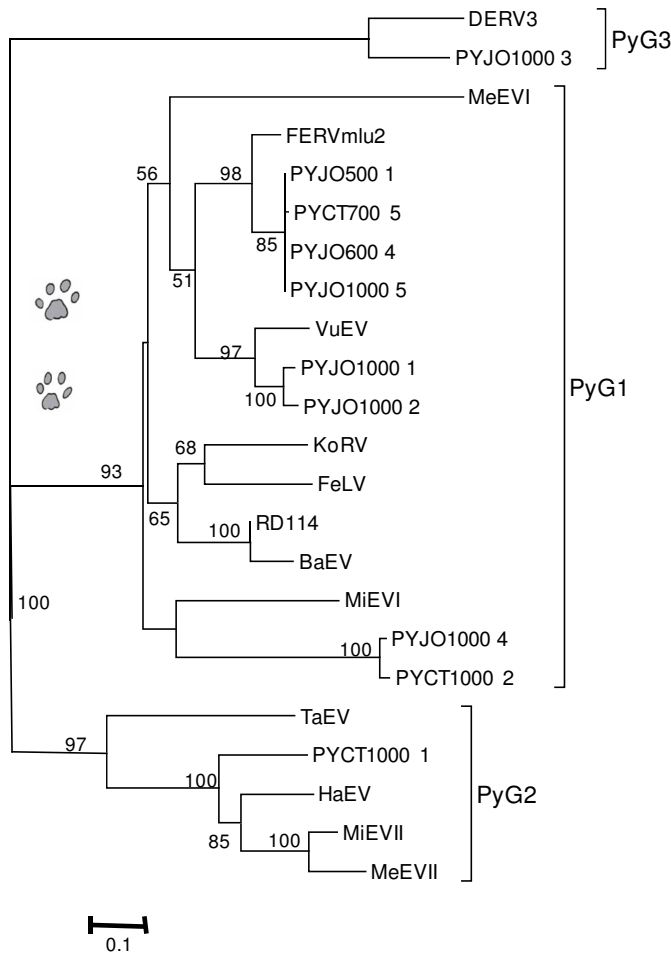
**Figura 13.** Filogenia ML baseada em 238 aminoácidos segundo modelo JTT+G, demonstrando relacionamentos entre os clones de *Leopardus colocolo*. Os clones LcoJo100018 e LcoJO120013 foram excuídos por apresentarem fragmento muito curto, mas agruparam-se dentro do clado representante de LcoG1 em uma NJ tree (análise não representada). Os clones de *L. colocolo* encontram-se descritos no Apêndice B e os demais ERVs, no Apêndice A. Os números próximos aos ramos representam os valores de *bootstrap* >50%.

### *Puma yagouaroundi*

Na espécie *Puma yagouaroundi* foram obtidos treze clones com sequências similares a ERVs. Destes, dez demonstraram similaridades com o gênero *Gammaretrovirus*. Todas as sequências mostraram-se defectivas, apresentando códons de parada e/ou inserções que provocaram mudança da fase de leitura.

Os dez clones *Gamma-like* (Tabela 1) apresentam-se divididos em três clados (PyG1 a PyG3), conforme demonstrado na Figura 14, demonstrando um padrão de diversidade similar ao descrito anteriormente para a maioria das espécies. No clado PyG1, obteve-se maior número de sequências (n=8). PyG2 e PyG3 são compostos de

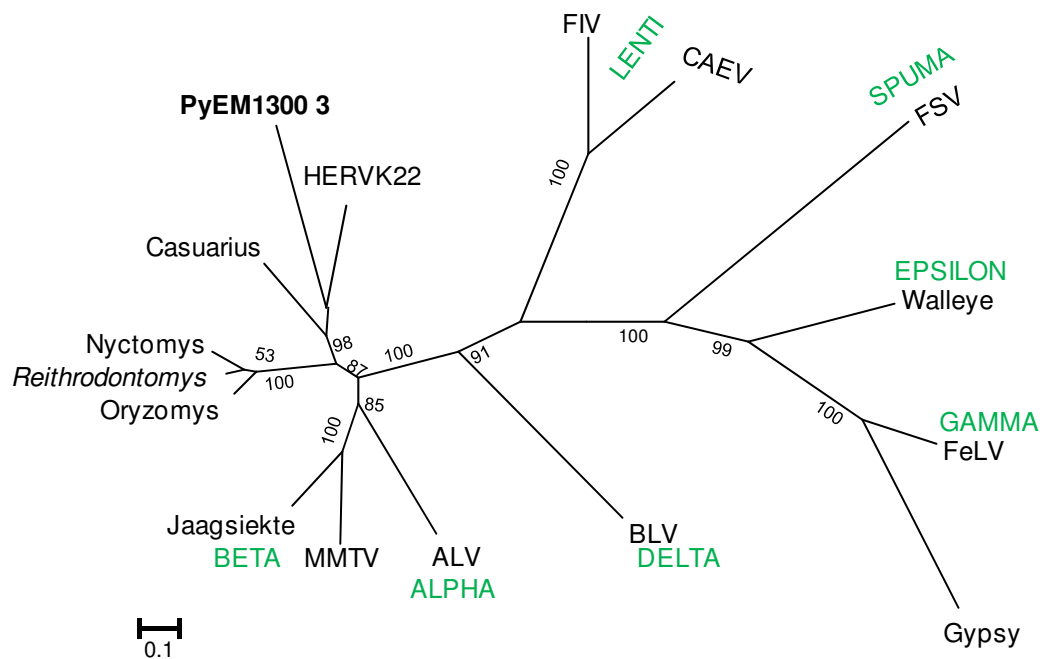
apenas uma sequência. Da mesma forma que nas demais espécies, o clado relacionado com FERVMul2 apresenta baixa divergência genética e será melhor analisado na Seção 5.3.3.



**Figura 14.** Filogenia ML baseada em 243 aminoácidos segundo o modelo WAG+G+F, demonstrando o relacionamento entre retroelementos *Gamma-like* de *Puma yagouaroundi* e ERVs descritos na literatura (Apêndice A).

O relacionamento filogenético dos clones PyEM13001, PyEM13003 e PyEM6002 com a família *Retroviridae* foi explorado através de buscas de similaridades com o programa BLAST e construção de filogenias cujos sete gêneros de retrovírus exógenos foram incluídos. Estes três clones são muito semelhantes entre si e apresentam similaridade com HERVK, que é classificado como *Betaretrovirus*. Duzentos e setenta e um pares de bases (posição 346 a 617) do clone PyEM13003 apresentaram uma

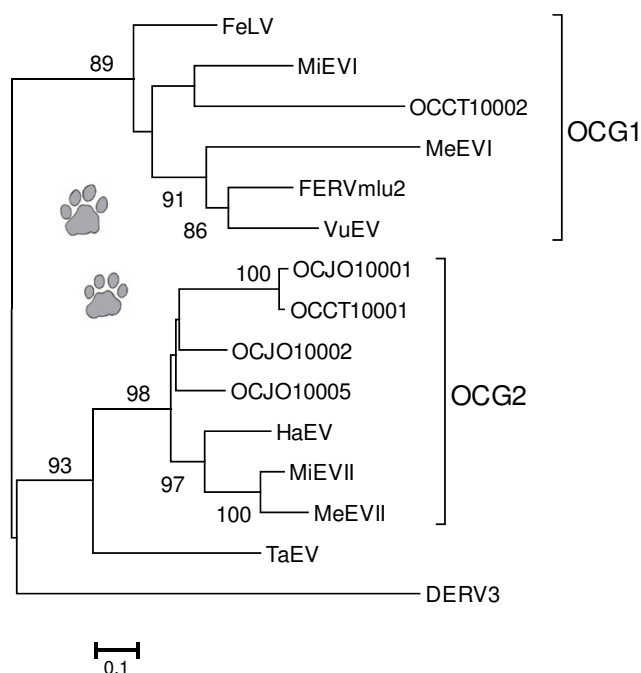
identidade de 69% ( $e\ value=7e-22$ ) com HERVK22 (GenBank: AC016955.15). As buscas realizadas no banco de dados nr do NCBI, através do programa BLASTn, também indicaram outros retrovírus endógenos similares a esses clones. Essas sequências foram incluídas na construção da filogenia demonstrada na Figura 15. Nessa filogenia, observa-se um grande grupo, contendo sequências de PyEM13003, HERVK22 do retrovírus endógeno, proveniente de uma ave (*Casuarious casuarious*), de ERVs pertencentes a roedores (*Nyctomys sumichrasti*, *Oryzomys palustris* e *Reithrodontomys fulvescens*) e dos retrovírus exógenos MMTV, Jaagsiekte e ALV. Esse grupo recebeu suporte estatístico alto na análise Bayesiana, bem como nas filogenias ML e NJ (não representadas). As sequências de ERVs desse grupo pertencem à classe II, portanto PyEM13003 e os dois clones similares (PyEM13001 e PyEM6002) fazem parte dessa classe.



**Figura 15.** Filogenia Bayesiana com base no alinhamento de aminoácidos demonstrado no Apêndice D e no modelo Rrev. Os retrovírus endógenos e exógenos apresentados na figura são listados no Apêndice A. Números próximos aos ramos representam a probabilidade posterior >50%.

### *Panthera onca*

Os cinco clones obtidos na espécie *Panthera onca* mostraram-se similares a *Gammaretrovirus*. Essas sequências agrupam-se em dois cladogramas distintos (denominados OCG1 e OCG2) (Figura 16). O clado OCG1 está representado por uma sequência apenas, OCCT10002. No grupo G3, representado pela sequência DERV3, não foram obtidos clones. A filogenia indica que o grupo OCG2 é o mais diverso, como demonstrado na maioria das espécies estudadas.



**Figura 16.** Filogenia ML baseada em um alinhamento de 238 aminoácidos segundo modelo JTT+G, demonstrando o relacionamento entre clones de *Panthera onca*. As sequências dos ERVs MeEVII, MiEVII, HaEV, VuEV, FERVmlu, MiEVI, MeEVI e FeLV são descritas no Apêndice A. Os números próximos aos ramos representam os valores de *bootstrap* >50%.

### 5.3.3. Caracterização dos clones do grupo *Myotis*

Neste estudo, foram encontradas algumas sequências que se agruparam com o retrovírus felino FERVmlu2, que, segundo a literatura, é muito similar a um ERV encontrado no morcego *Myotis lucifugus*. Essas sequências foram analisadas, de modo sucinto, nas descrições das filogenias intraespecíficas.

A fim de investigar se essas sequências estariam presentes em outros organismos, foi realizada uma busca por sequências homólogas (BLASTn) no banco de dados WGS, utilizando-se como *query* uma sequência consenso de 608 pb construída no programa Bioedit 7.0.5 (Hall, 1999). Para montagem desta sequência, foram escolhidas 13 sequências (Pco, n = 7; Lwi, n = 1; Lg, n = 1; Py, n = 4) que apresentaram maior número de nucleotídeos, considerando que as demais sequências recuperadas nessa linhagem possuíam fragmentos pequenos (aproximadamente 300 pb). Obtiveram-se 283 resultados (*hits*) com *F. catus* (*e value* < 9e-111) com identidade variando entre 65% e 87%. O segundo grupo de sequências com identidade mais alta com a *query* (73-78% de identidade) é representado pela espécie *Myotis lucifugus*, com 40 resultados (*e value* < 9e-86). Escores significativos também foram encontrados em uma variada lista de mamíferos placentários e em uma sequência de *Trypanosoma congolense*, mas com similaridades menores. As sequências do banco com maior escore encontradas em cada espécie estão listadas na Tabela 3.

Em uma nova estratégia de busca por similaridades no banco de dados WGS, foram utilizados 329 aminoácidos (fragmento PR-RT) da sequência FERVmlu2 e o programa tBLASTn. Foram recuperados 133 resultados de *Felis catus*, 17 de *Myotis lucifugus*, 2 de *Heterocephalus glaber* e 1 de *Tarsius syrichta*. A sequência com maior escore de *Felis catus* (IACBE01338934.1) apresentou 92% de identidade com FERVmlu2, a de *Myotis lucifugus* (AAPE02049384.1), 85%, e as de *Heterocephalus glaber* (AFSB01147810.1) e *Tarsius syrichta* (ABRT010822206.1), 70%.

Com o propósito de estabelecer relacionamentos entre os ERVs relacionados com FERVmlu 1 e 2 e aqueles encontrados no GenBank, foram construídas filogenias seguindo as seguintes estratégias:

- 1) Utilização de todas as sequências obtidas neste estudo relacionadas a FERVmlu 1 e 2, acrescentando algumas sequências descritas na Tabela 3, considerando-se estes critérios: número de pb e saturação (verificada no programa Dambe); filogenia representada na Figura 17.
- 2) Seleção das sequências com maior número de caracteres e análise de seus respectivos aminoácidos; filogenia representada na Figura 19.

**Tabela 3.** Lista das espécies com sequências similares à linhagem *Myotis*, obtidas através de buscas realizadas com o programa BLASTn no banco de dados WGS do GenBank. Na tabela, estão representadas as sequências do banco (*hit*) com maior escore para cada espécie. A classificação representa a ordem a qual pertence cada espécie.

<b>Espécie</b>	<b>Score</b>	<b>Identidade</b>	<b>%</b>	<b>E value</b>	<b>GenBank</b>	<b>Fragmento <i>query</i> (pb)</b>	<b>Classificação</b>
<i>Felis catus</i>	830	504/576	87	0	AANG02121168.1	32-607	Carnivora
<i>Myotis Lucifungus</i>	596	455/581	78	4 e-166	AAPE02021020.1	27-607	Chiroptera
<i>Otolemur garnettii</i>	284	344/514	67	4.00E-64	AAQR03072652.1	95-608	Primates
<i>Heterocephalus glaber</i>	210	320/486	66	6.00E-50	AFSB01151231.1	72-552	Rodentia
<i>Tarsilus syrichta</i>	201	274/413	66	3.00E-47	ABRT010584641.1	192-604	Primates
<i>Macaca fascicularis</i>	178	247/364	68	3.00E-40	CAEC01092414.1	246-607	Primates
<i>Macaca mulata</i>	178	246/365	67	3.00E-40	AANU01293367.1	246-607	Primates
<i>Rattus novergicus</i>	174	244/357	68	4.00E-39	AABR05039834.1	196-546	Rodentia
<i>Dipodomys ordii</i>	170	239/361	66	5.00E-38	ABRO01269960.1	192-552	Rodentia
<i>Spermophilus tridecemlineatus</i>	169	258/383	67	2.00E-37	AGTP01040106.1	192-569	Rodentia
<i>Mustela putorius furo</i>	169	311/480	65	2.00E-37	AEYP01094289.1	80-550	Carnivora
<i>Pan troglodytes</i>	167	243/364	67	6.00E-37	AACZ03013383.1	246-607	Primates
<i>Canis lupus familiaris</i>	158	315/487	65	3.00E-34	AAEX03006883.1	72-554	Carnivora
<i>Gorilla gorilla</i>	158	241/364	66	3.00E-34	CABD02405173.1	246-607	Primates

<b>Espécie</b>	<b>Score</b>	<b>Identidade</b>	<b>%</b>	<b>E value</b>	<b>GenBank</b>	<b>Fragmento <i>query</i> (pb)</b>	<b>Classificação</b>
<i>Cavia porcellus</i>	158	264/404	65	3.00E-34	AAKN02057207.1	139-541	Rodentia
<i>Mus musculus</i>	158	276/425	65	3.00E-34	AAHY01075839.1	109-532	Rodentia
<i>Oryctolagus cuniculus</i>	156	244/362	67	1.00E-33	AAGW02033344.1	198-554	Lagomorpha
<i>Erinaceus europaeus</i>	150	242/367	66	5.00E-32	AANN01173148.1	190-554	Erinaceomorpha
<i>Cricetulus griséus</i>	149	241/362	67	2.00E-31	AFTD01071971.1	190-546	Rodentia
<i>Microcebus murinus</i>	149	292/455	64	2.00E-31	ABDC01405982.1	105-554	Primates
<i>Procavia capensis</i>	143	247/379	65	7.00E-30	ABRQ01458093.1	181-555	Hyracoidea
<i>Tupaia belangeri</i>	138	238/364	65	3.00E-28	AAPY01301964.1	190-552	Scandentia
<i>Trypanosoma congolense</i>	136	271/425	64	1.00E-27	CAEQ01000580.1	109-532	Protista/Trypanosomatida
<i>Tursiops truncatus</i>	136	235/364	65	1.00E-27	ABRN01427179.1	192-554	Cetacea
<i>Pteropus vampyrus</i>	131	225/344	65	4.00E-26	ABRP01315765.1	190-532	Chiroptera

(Tabela 3. Continuação)



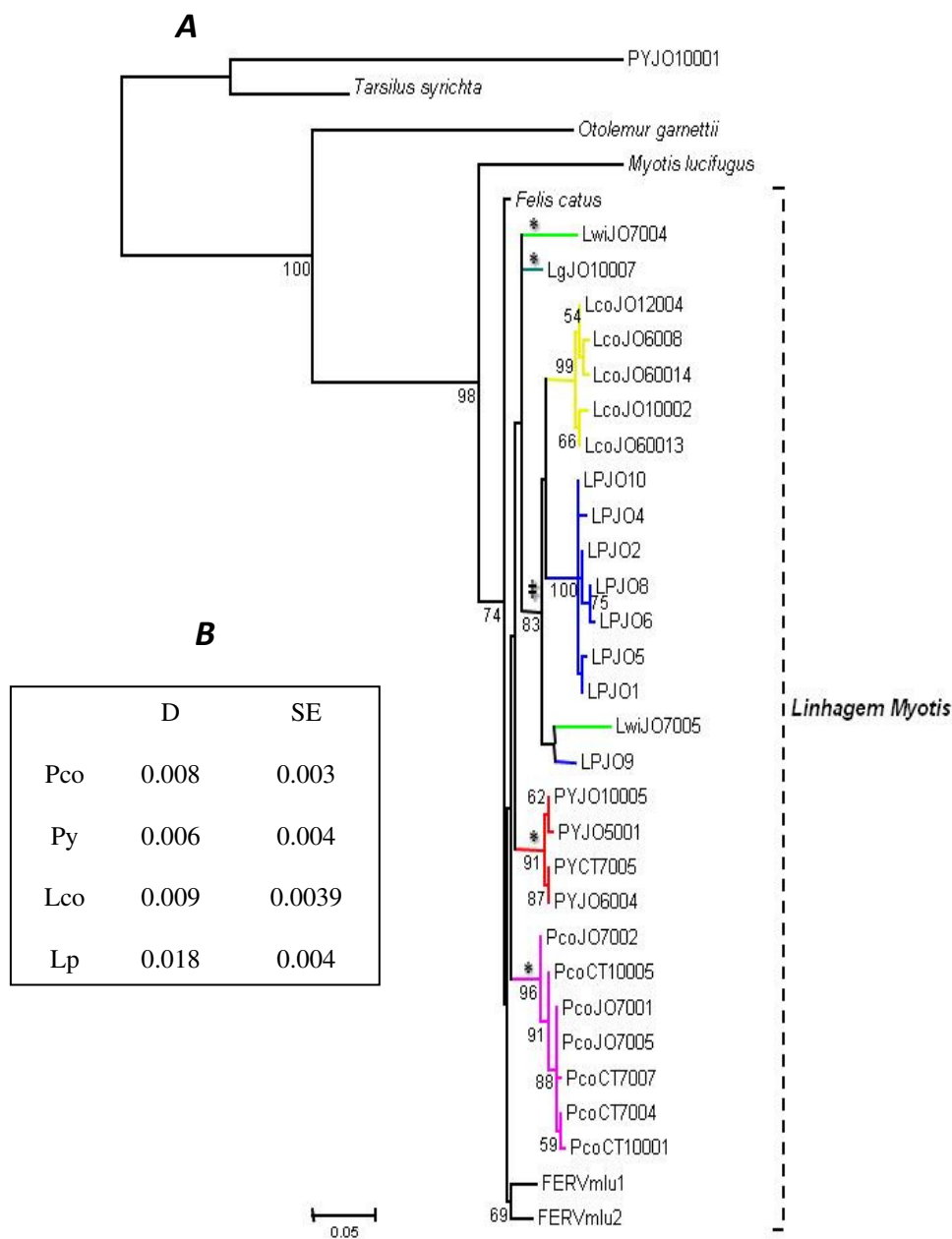
A filogenia da Figura 17A corrobora os resultados encontrados nas filogenias intraespecíficas e a similaridade sugerida pela busca no GenBank (Tabela 3), agrupando com alto suporte de *bootstrap* (98%) a linhagem de sequências provenientes de felídeos (denominada neste estudo como linhagem *Myotis*) com a sequência de *Myotis lucifugus*. O clone PyJO10001, pertencente a um clado distinto da linhagem *Myotis* (Figura 14), agrupou-se com *Tarsilus syrichta*, que apresentou similaridade de 66% em relação à sequência consenso, utilizada como *query* (Tabela 3).

Na linhagem *Myotis*, o esboço da filogenia sugere uma expansão recente. Observa-se a presença de um clado composto por sequências de *L. colocolo*, *L. pardalis*, *L. wiedii* e *L. geoffroyi*, ou seja, formado exclusivamente por espécies da linhagem da Jaguatirica (Johnson *et al.* 2006), um grupo monofilético de sequências de *Puma yagouaroundi* e outro de *P. concolor*. As duas sequências de *L. wiedii*, como descrito na Seção 5.3.2, apresentam uma divergência mais profunda, não acompanhando o padrão das outras espécies. A alta similaridade intraespecífica (Figura 17B) reflete o maior número de cladogramas monofiléticos (*P. concolor*, *P. yagouaroundi* e *L. colocolo*) e menor número de parafiléticos (*L. pardalis*), demonstrados na filogenia.

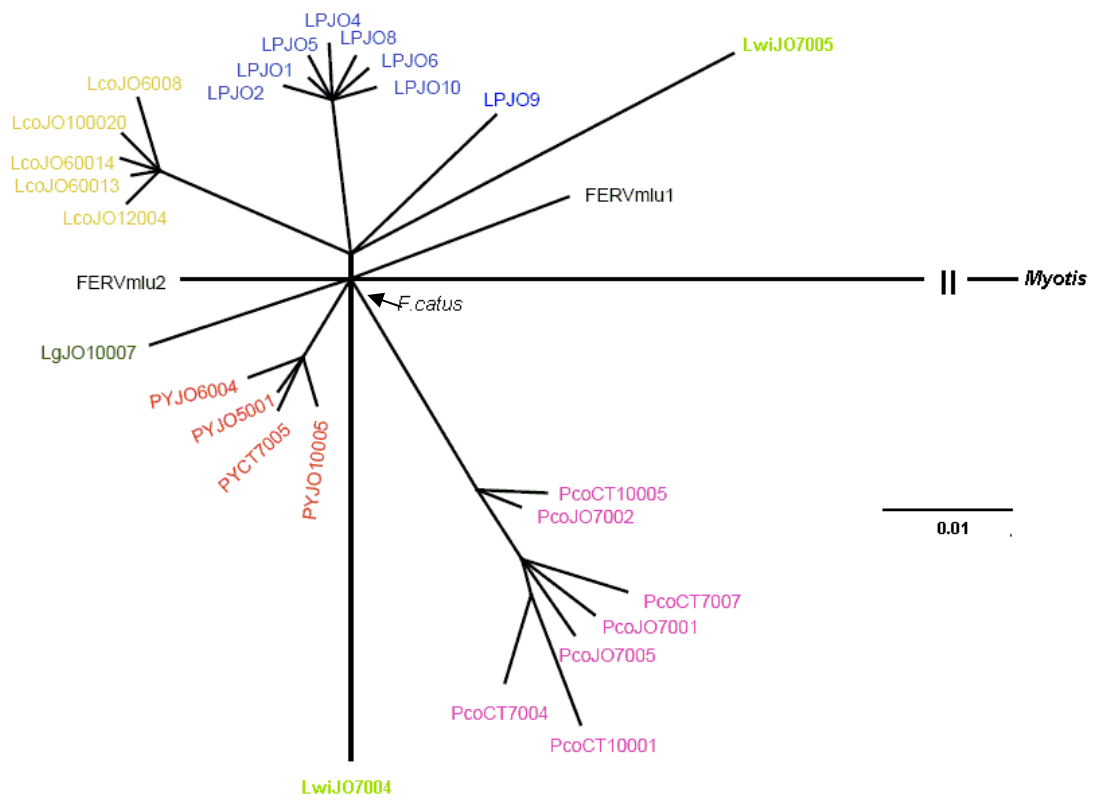
A filogenia NJ (não representada) foi muito semelhante à ML, reforçando todos os agrupamentos, até mesmo aqueles com valores de confiança abaixo de 50%. A *Network* (Figura 18) também corrobora todos os resultados descritos e o seu esboço em forma de estrela demonstra a baixa resolução dos relacionamentos interespecíficos destes retroelementos neste conjunto de dados.

Quando se observa o alinhamento das sequências da linhagem *Myotis*, constata-se que essas sequências são divididas em dois grupos no que diz respeito ao tamanho do fragmento. Esses grupos estão representados na Figura 17A. Um grupo (clado formado por *L. pardalis*, *L. colocolo* e LwiJO7005) contém sequências de aproximadamente 330 pb, que apresentam deleções da protease e parte da transcriptase reversa.

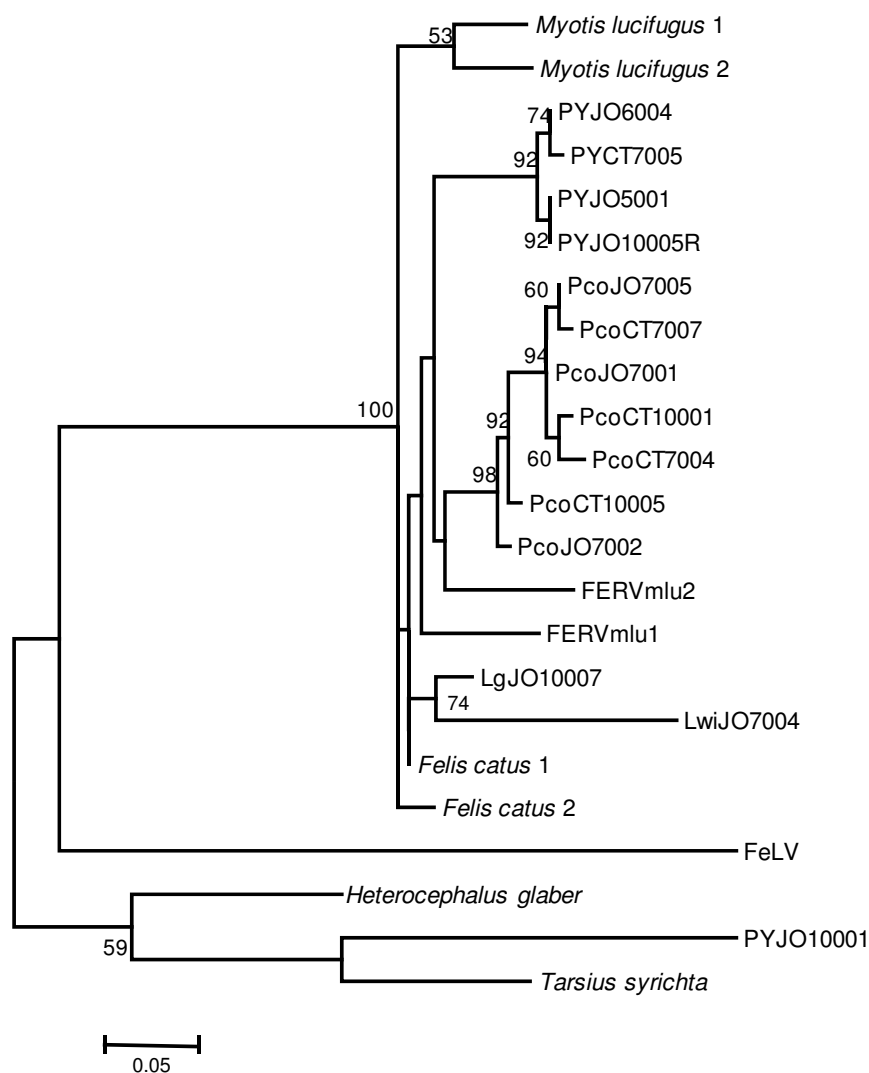
O outro grupo (clado formado por *P. concolor*, *P. yagouarondi*, LwiJO7004 e LgJO10007) é composto por sequências de 600 pb e contém o *motif* da protease VLVDTG no início do fragmento, deleção de todo o resto da protease e de aproximadamente 100 pb da região da RT. A sequência de *F. catus* obtida no GenBank não apresentou as deleções descritas acima.



**Figura 17.A.** Filogenia ML com 325 pb (do 2° ao 5° domínio da RT) segundo modelo GTR+G. O alinhamento continha 179 nucleotídeos variáveis sendo 108 deles informativos para parcimônia. Os *taxa* em itálico estão descritos na Tabela 3, clones no Apêndice B e FERVmlu1 e 2 no Apêndice A. Valores de *bootstrap* > 50% são apresentados junto aos ramos. ‡ = clado com 330 pb; \* = clado com 600 pb. B. Distância genética p (D) intraespecífica das sequências apresentadas na Figura 17A; SE= erro padrão.



**Figura 18.** *Split decomposition network* da linhagem *Myotis* estimada com 325 pb da região da RT segundo o modelo GTR .



**Figura 19.** Filogenia ML construída com 192 aminoácidos (1° ao 5° domínios da RT) segundo o modelo JTT+G. Os clones Lg, Lwi, Pco e Py estão descritos no Apêndice B. *Myotis lucifugus* 1 *Felis catus* 1 (correspondem aos *taxa* descritos na Tabela 3); *Myotis lucifugus* 2 (GenBank n° AAPE02049384.1) *Felis catus* 2 (GenBank n° IACBE01338934.1), *Heterocephalus glaber* (GenBank n° AFSB01147810.1) e *Tarsius syrichta* (GenBank n° ABRT010822206.1) conforme no texto; FERVmlu1 e FERVmlu2 (Apêndice A).

A filogenia baseada em aminoácidos, demonstrada na Figura 19, recuperou um padrão similar ao obtido com a de nucleotídeos com as sequências de *Puma yagouaroundi* e *P. concolor* dispostas em clados monofiléticos (ver a Figura 17A). As sequências LgJO10007 e LwiJO7004 formaram um grupo isolado. A principal diferença

entre as filogenias é o relacionamento dos FERVmlu1 e 2 dentro da linhagem *Myotis*, no entanto com baixo suporte estatístico.

Um fato interessante foi a recuperação de um fragmento de 423 pb teoricamente inserido no genoma de *Trypanosoma congolense* apresentando 64% de similaridade com a sequência consenso da linhagem *Myotis* (Tabela 3). Utilizou-se esse fragmento como *query* em comparação com o banco de dados WGS, através do programa tBLASTx. Nessa consulta, foram recuperadas três sequências de *Trypanosoma congolense* com 97-100% de identidade (GenBank CAEQ01000580.1, CAEQ01001954.1 e CAEQ01001363.1) que apresentavam regiões similares aos genes *gag* e *pol*. A sequência de aminoácidos de *Trypanosoma congolense* CAEQ01000580 agrupou-se com o gênero *Gammaretrovirus*, mas fora da linhagem *Myotis* (filogenia não representada). Através de uma busca (BLAST) realizada em 08 de dezembro de 2011 no banco de dados UniprotKB ([www.uniprot.org](http://www.uniprot.org)), observou-se que essa sequência não havia sido revisada. Portanto, qualquer inferência sobre o resultado relatado acima deverá ser analisada com muita cautela.

#### 5.3.4. Análises interespecíficas

Na Figura 20, observam-se as relações interespecíficas entre linhagens de ERVs *Gamma-like* encontrados nas oito espécies de felídeos estudadas. Nessa filogenia, observa-se a divisão dos ERVs *Gamma-like* em três grandes grupos (G1 a G3) reprisando os resultados obtidos nas análises intraespecíficas.

A distância genética (distância p) entre as sequências de felídeos dentro do grupo G1 ( $d=0,312$ ;  $SE=0,009$ ), G2 ( $d=0,181$ ;  $SE=0,005$ ), G3 ( $d=0,083$ ;  $SE=0,006$ ) demonstra que a variação dentro do grupo G1 é bem maior do que a dos demais grupos, refletindo o maior grau de dispersão das sequências observados na filogenia com formação de subgrupos mais diferenciados entre si.

As sequências pertencentes ao grupo G1 são provenientes das espécies *P. concolor*, *P. yagouaroundi*, *L. colocolo*, *L. geoffroyi*, *L. wiedii* e *F. catus*. A espécie *Panthera onca* está representada nesse grupo pela sequência OCCT10002 (Figura 16), mas foi excluída da análise por apresentar apenas a região da protease e uma pequena porção da RT. A espécie *L. pardalis* também possui representantes nesse grupo, que foram excluídos da análise por apresentarem sequências muito curtas. Essas sequências

pertencem à linhagem *Myotis* descrita anteriormente (Figuras 17 e 18). Portanto, com exceção de *L. tigrinus* todas as outras espécies estudadas estão representadas nesse clado. Duas linhagens principais ficam evidentes nesse grupo: uma relacionada com MiEVI, nomeada G1A por motivos didáticos, e outra com FeLV, denominada G1B (Figura 20).

As sequências do clado G1A agrupam-se a MiEVI com baixo suporte estatístico e, na filogenia demonstrada na Figura 21, essas sequências formam um clado basal e isolado. O longo ramo que separa esse grupo de sequências do nó interno da árvore (Figuras 20 e 21) sugere ter havido extinção de linhagens proximamente relacionadas, ou, a existência de sequências que não foram amostradas.

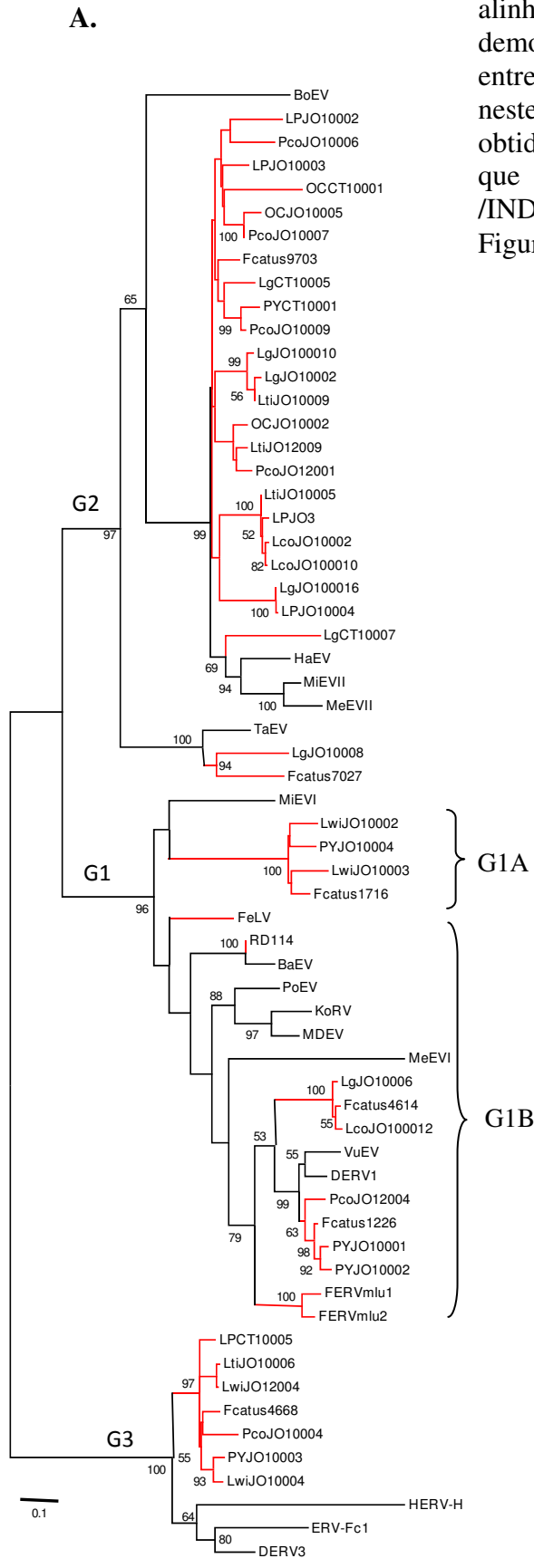
O clado G1B apresenta três linhagens de ERVs provenientes de felídeos bastante distintas: 1) representada por FeLV, 2) representada por RD114, e 3) representada por um grupo contendo algumas das novas sequências dos ERVs, duas sequências do gato doméstico e os ERVs FERVmlu 1 e 2 (linhagem *Myotis* descrita anteriormente). Então, ao contrário de FeLV e RD114, que estão enclausuradas no genoma de felídeos do gênero *Felis*, as sequências dessa última linhagem são mais difundidas na família Felidae.

O Grupo G2 está dividido em dois subgrupos, um deles composto por uma sequência de *L. geoffroyi* e *F. catus* está proximamente relacionado ao ERV TaEV, proveniente de um quiróptero da espécie *Tadarida brasiliensis*. No outro subgrupo, encontram-se presentes *F. catus* e todas as espécies de felídeos estudadas, exceto *L. wiedii*. Esse subgrupo divide-se em dois grupos menores: 1) um contém a sequência LgCT10007, que demonstra similaridades com os ERVs dos carnívoros HaEV (*Halichoerus grypus*), MeEVII (*Meles meles*) e MiEVII (*Mustela vison*); 2) o outro é formado por vários clados pequenos que apresentam ramos internos muito curtos (forma de estrela), sugerindo uma expansão simultânea.

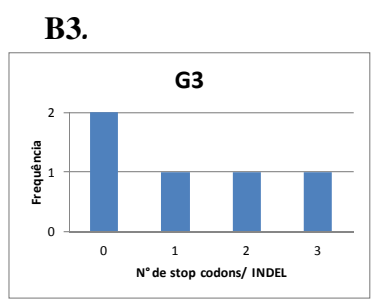
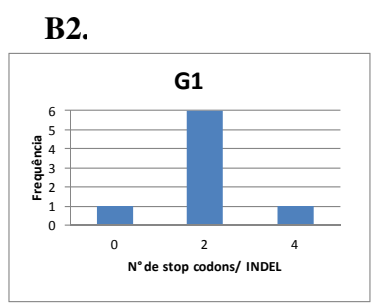
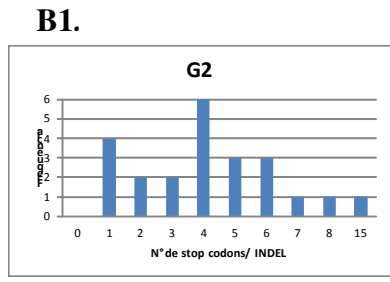
O grupo G3 é composto por sequências oriundas de quase todas as espécies estudadas exceto *P. onca* e *L. colocolo*. O gato doméstico também está presente nesse grupo, o qual se mostrou menos diversificado, com sequências constituindo um único clado. Além disso, foram amplificadas apenas uma ou, muito raramente, duas sequências por espécie, sugerindo que elas estão presentes em pequeno número de

cópias. O grupo G3 relaciona-se com DERV3, um ERV proveniente do cão, e com uma família de ERVs conhecida como HERV F/H, presente no genoma de primatas.

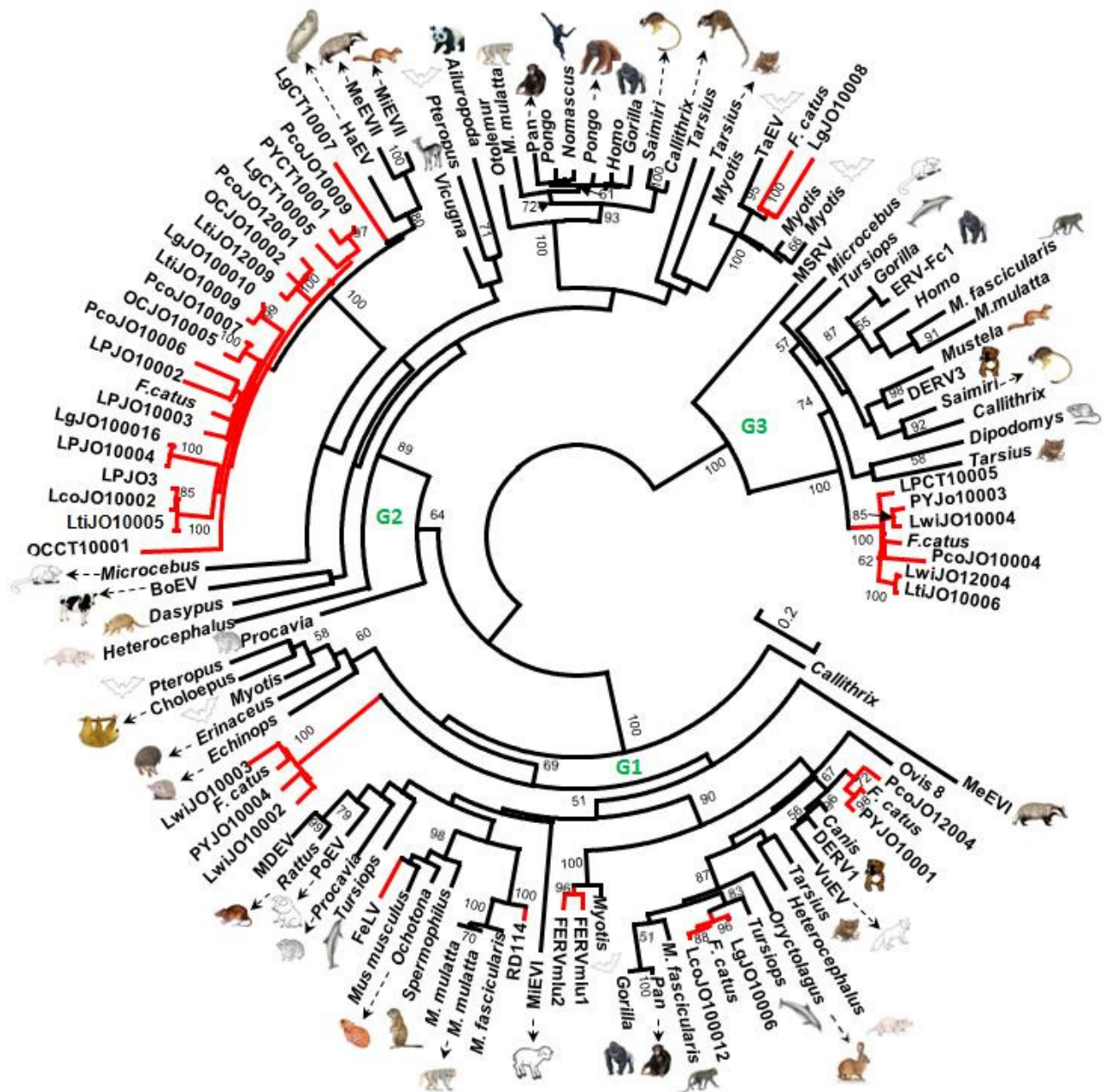
Em relação à funcionalidade inferida pela presença de códons de parada e/ou de inserções/deleções que causem mudanças na fase de leitura, observou-se um perfil diferente entre os três grupos (Figura 20B). O G2 tem todas as suas sequências defectivas, apresentando um valor de mediana de quatro *stopcodons/indel*, provavelmente representando antigas linhagens virais; o G1 apresenta a sequência LgJO10006 com ORF parcial, correspondente à região sequenciada aberta (funcional) e um valor de mediana de dois *stopcodons/indel*; o G3 apresenta as sequências LtiJO10006 e LwiJO12004 com ORF parcial aberta e valor de mediana de 1,5 *stopcodons/indel*, sugerindo que muitos desses ERVs teriam sido ativos em um passado recente.



**Figura 20 A.** Filogenia ML com base no alinhamento de 240 aminoácidos (Figura 4), demonstrando os relacionamentos filogenéticos entre as sequências dos ERVs *Gamma-like* obtidos neste estudo (Apêndice B) e sequências de ERVs obtidas no GenBank (Apêndice A). **B1-B3.** Gráficos que demonstram a frequência de *stop codon* /INDEL nas sequências dos clones analisados na Figura 20 A nos grupos G1-G3.







**Figura 21.** Reconstrução filogenética por máxima verossimilhança com base em um alinhamento de 228 aminoácidos (68 da região PR e 160 da região da RT) segundo o modelo JTT+G. As sequências obtidas através do programa tBLASTn estão descritas no Apêndice E, de acordo com o grupo G1, G2 e G3, indicados em verde. As sequências dos demais ERVs estão listadas no Apêndice A. Os ramos de origem Felidae estão destacados em vermelho.

Para compreender melhor a evolução dos ERVs *Gamma-like* obtidos neste estudo, foi investigado o comportamento desses retroelementos, quando comparados com sequências obtidas no banco de dados WGS de todos os vertebrados disponíveis no

GenBank. As buscas foram realizadas com o programa tBLASTn no início de dezembro de 2011 (Apêndice E).

A filogenia da Figura 21 consiste em um exemplo claro de que ERVs relacionados (similares) com os dos felídeos teriam infectado pelo menos 11 ordens de mamíferos. Os três grupos descritos acima (G1-G3) são mantidos, bem como os seus subgrupos, sendo que a maioria deles apresenta sequências de outras ordens/famílias de mamíferos agregadas. Dentro desses subgrupos, observa-se, em muitos casos, que existem vírus isolados em grupos de hospedeiros ocorrendo uma concordância entre as filogenias destes. A título de exemplo, no grupo G2, observou-se: 1) que a sequência LgCT10007 agrupou-se com carnívoros de acordo com seus relacionamentos filogenéticos e, 2) que o clado dos primatas demonstrou uma clara divisão entre primatas do novo e do velho mundo.

Existem casos em que as sequências de felídeos estão isoladas em subgrupos menores, agrupando-se com *taxa* filogeneticamente mais distantes como, por exemplo, o clado da TaEV no grupo G2 e os cladus FERVmlu1, RD114 e FeLV no grupo G1. Essa observação sugere infecção de linhagens de retrovírus filogeneticamente muito próximas entre espécies de mamíferos evolutivamente distantes.

Foi avaliado se houve seleção em um fragmento de 456 pb da região da RT em um grupo de sequências de clones obtidas neste estudo somada às de *Felis catus* adquiridas no GenBank, conforme a filogenia da Figura 20. Na análise de seleção assumiu-se a premissa de que os grupos avaliados eram compostos por sequências ortólogas, com base na baixa divergência genética observada entre elas, e na resolução filogenética demonstrada na Figura 20: no grupo G2 criou-se um subgrupo formado por todas as sequências provenientes de felídeos exceto LgCT10007, LgJO10008 e *F. catus* 7027; O Grupo G3 é composto por todas as sequências correspondentes ao clado G3.

Observaram-se os seguintes valores de  $\omega$  (DN/DS): no subgrupo de G2,  $\omega = 0,744$  (95% CI: 0,649; 0,847); e, no grupo G3,  $\omega = 0,377$  (95% CI: 0,266; 0,516). Estes resultados sugerem seleção purificadora nos dois grupos analisados, apresentando intervalos de confiança distantes de 1, o que indicaria uma situação neutra. Entretanto, os resultados para o grupo G3 devem ser avaliados com cautela devido ao pequeno número de sequências analisados ( $n=6$ ). Em cada alinhamento, foi sugerido apenas um evento de recombinação (sítio 293, no subgrupo de G2; sítio 384 no G3) suportado

pelas estatísticas AIC e AICc e nenhum evento pela estatística BIC, que é mais conservadora. Esse baixo número de sítios recombinantes encontrados nos alinhamentos foi proporcional ao limiar encontrado em simulações realizadas por Anisimova *et al.* (2003), nas quais a acurácia dos resultados não foi afetada.

#### 5.4. Recombinação

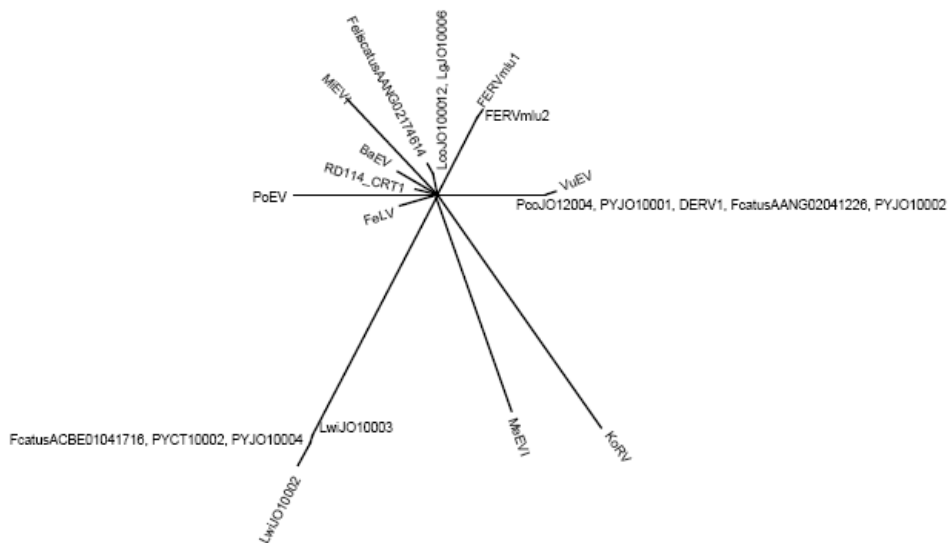
Filogenias construídas com método de máxima verossimilhança cujos fragmentos da protease (Apêndice F) e transcriptase reversa (Apêndice G) foram analisados separadamente demonstraram a presença característica dos três grupos (G1-G3). Não ocorreu migração de sequências entre eles, de modo que nenhum evento óbvio de recombinação foi demonstrado entre os grupos. Dentro dos grupos observam-se algumas incongruências nas filogenias com rearranjos entre alguns *taxa*. Para observar com mais detalhes a evolução dentro de cada grupo, foram construídas *networks* por meio do método *Split decomposition* (ver a Seção 4, Métodos), sendo que os fragmentos da protease e transcriptase reversa foram analisados separadamente (Figuras 22, 23 e 24). O alinhamento de nucleotídeos utilizado para construir as *networks* foi baseado no alinhamento de proteína, preservando-se a fase de leitura. As *networks* foram estimadas com a distância Kimura 3ST. Devido à facilidade de se utilizar o programa SplitsTree 4, outras distâncias também foram avaliadas, usando esse programa, as quais produziram resultados semelhantes.

As *networks* não demonstram evidência de evolução reticulada nos grupos G1 e G2, compondo uma *network* em forma de estrela, a qual reflete os baixos suportes estatísticos obtidos nas filogenias (Figuras 22 e 23). Também não foi observada incongruência filogenética óbvia, que pudesse sugerir recombinação, quando comparadas as *networks* produzidas com os fragmentos de PR e RT. O grupo G3 demonstrou uma *split network* com aparência reticulada em ambos os fragmentos (Figura 24), entretanto, o teste *phi* não encontrou evidência de recombinação nos segmentos de PR e RT analisados ( $p > 0,7$ ). A sequência LpCT10005 foi excluída da amostra (Figura 24) por apresentar fragmento curto na região da protease. Entretanto, quando incluída na análise, os resultados (*networks* e teste *phi*) foram similares. Esses resultados não indicam, necessariamente, que a homoplasia encontrada no grupo G3 seja consequência de mutações recorrentes. Como este é um grupo muito recente, a

sequência ancestral apresenta características muito similares às das sequências existentes. O índice de homoplasia  $\phi$  é uma medida de similaridade entre sítios proximamente ligados. Em um modelo finito de recombinação, sítios distantes tendem a ser geneologicamente menos correlacionados do que sítios próximos. Em um modelo nulo de recombinação (mutação recorrente), ao contrário, todos os sítios dividem a mesma história. Portanto, o teste pode não ter diferenciado os dois modelos descritos acima. Outro fato a ser considerado é que o número de amostras ( $n=7$ ) pode não ter sido suficiente para obter significância no teste através de uma distribuição normal.

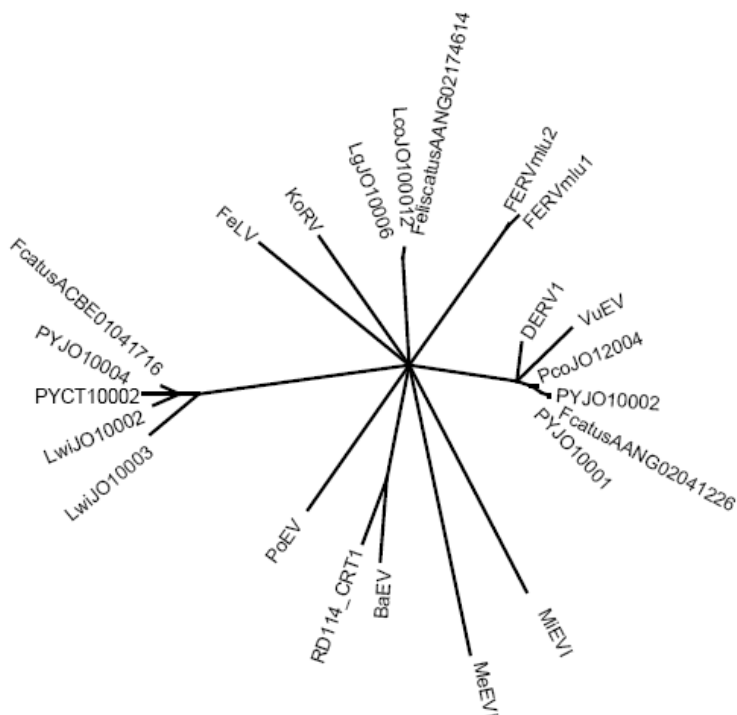
## G1-PR

H<sub>0,01</sub>



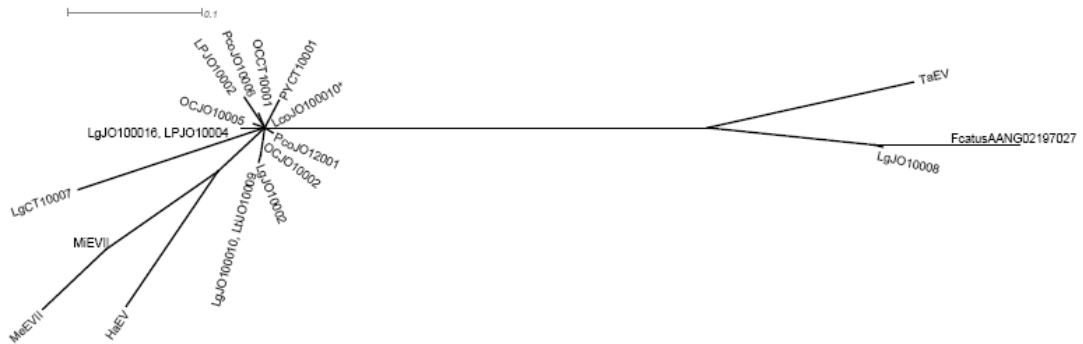
## G1-RT

H<sub>0,01</sub>

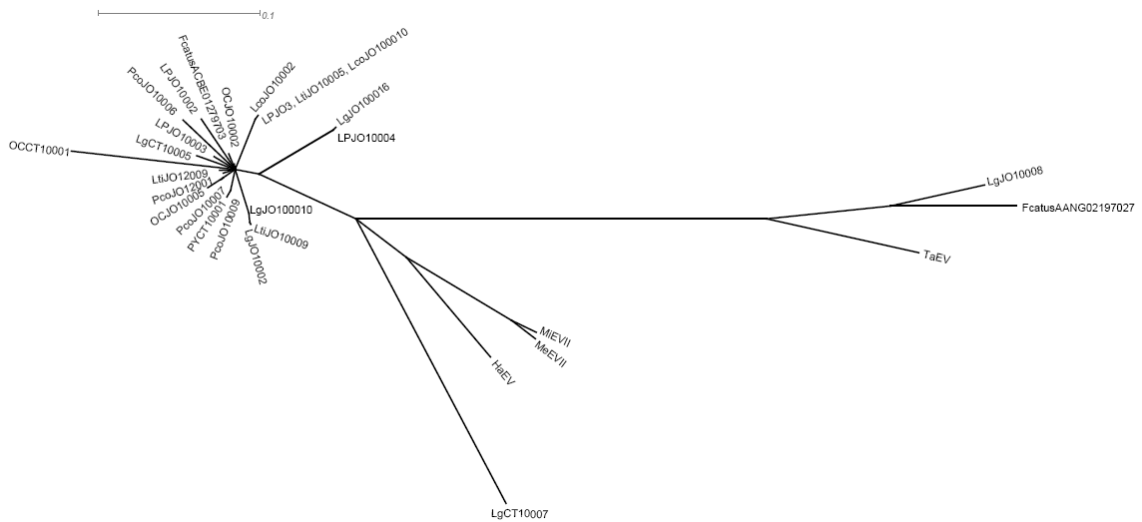


**Figura 22.** *Split decomposition network* de 23 seqüências relacionadas ao grupo G1, baseado em um alinhamento de 267 pb do gene PR (G1-PR) e 686pb do gene RT (G1-RT).

## G2-PR

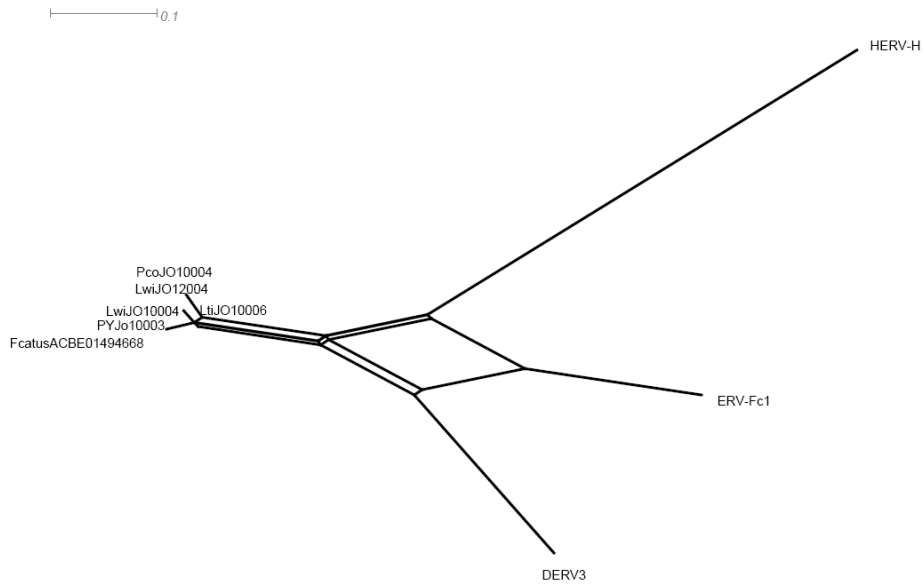


## G2-RT

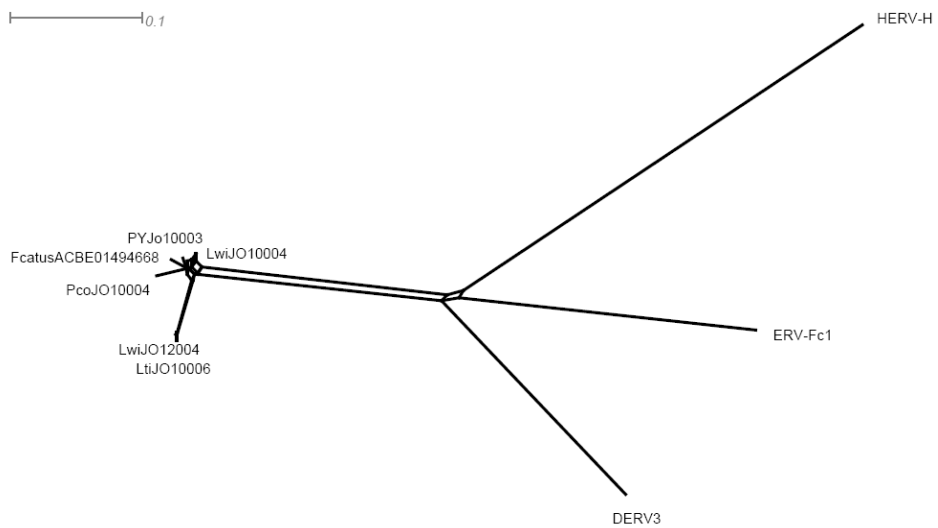


**Figura 23.** Split decomposition network de 29 sequências relacionadas ao grupo G2, baseada em um alinhamento de 316 pb do gene PR (G2-PR) e 699 pb do gene RT (G2-RT). A sequência LcoJO100010 da Figura G2-PR é um representante do seguinte conjunto de sequências (LtiJO10005, LpJO3, LcoJO10002, *F. catus* ACBE01279703, LtiJO12009, LgCT10005, PcoJO10007, PcoJO10009, LpJO10003).

### G3-PR



### G3- RT



**Figura 24.** *Split decomposition network* de nove sequências relacionadas ao grupo G3, baseado em um alinhamento de 255 pb do gene da protease (G3-PR) e 570 pb do gene da transcriptase reversa (G3-RT).

## 6. DISCUSSÃO

### 6.1. Diversidade e funcionalidade de ERVs no genoma dos felídeos

A grande maioria dos ERVs obtida neste estudo é defectiva, apresentando mutações e *indels* que provocam a interrupção ou mudança da fase de leitura. Então, aparentemente, não existe nenhuma pressão para manter a integridade desses retrovírus. Por consequência, o acúmulo de mutações a cada geração é irreversível e, geralmente, toda a linhagem torna-se defectiva, um fenômeno conhecido como Catraca de Muller (Muller, 1964). Os ERVs que não podem mais replicar, com o passar do tempo, provavelmente, se extinguem. Os retrovírus exógenos se diferenciam dos endógenos nesse aspecto, porque eles mantêm seu *fitness* por meio de alta taxa de mutação e seleção da variante mais adaptada (Grifford & Tristem, 2003).

Uma das características mais interessantes encontrada neste estudo foi a predominância de sequências *Gamma-like*. Os *Gammaretrovirus* são ERVs, também dominantes em chimpanzés, humanos, gambás e cães (Blomberg *et al.*, 2011), além de serem difundidos entre os vertebrados (Gifford & Tristem, 2003).

A observação da grande diversidade de ERVs similares a *Gammaretrovirus* nas espécies estudadas leva a questionamentos sobre a sua dominância. Alguns autores atribuem o sucesso da endogenização de um ERV à presença de receptores específicos utilizados pelo vírus, particularmente na célula germinativa (Denner, 2010). Além disso, a exploração do meio endógeno por um vírus impõe certas resistências por parte do hospedeiro, as quais podem variar de espécie para espécie (Holmes, 2010), por exemplo, o desenvolvimento de mecanismos de proteção e defesa (Münk *et al.*, 2008). Consequentemente, a manutenção de características comuns permitiria aos retrovírus explorar grupos taxonômicos distintos. Outra característica igualmente importante é a recombinação, porque permite que os retrovírus explorem novos nichos taxonômicos através da aquisição de novas informações genéticas.

Curiosamente, o espelhamento dos *Gammaretrovirus* endógenos nas diferentes linhagens de felídeos, como demonstrado neste estudo, não corresponde ao atual padrão de distribuição do seu representante exógeno. O FeLV infecta esporadicamente felídeos silvestres, apresentando alta prevalência nos gatos domésticos, provavelmente, devido à



estrutura populacional e mudanças de comportamento impostos pela domesticação. As sequências endógenas de FeLV são encontradas apenas em felídeos do gênero *Felis*, e alguns estudos sugerem que FeLV está circulando nas populações das espécies desse gênero há algum tempo (Polani *et al.*, 2010). FeLV ocorre na natureza como uma complexa família de genomas proximamente relacionados, e a sua diversidade é gerada pelas mutações ocorridas durante a transcrição reversa e a recombinação com FeLV endógenos (Ahmad & Levy, 2010). Assim, o FeLV endógeno poderia ser considerado um estoque de diversidade, como talvez sejam os outros elementos *Gamma-like*, cujos exógenos teriam sido extintos.

## 6.2. Relacionamento dos ERVs com seus hospedeiros

Em gatos domésticos os retrovírus exógenos FIV e FeLV causam doenças crônicas com um longo período de latência, influenciando a longevidade do indivíduo. Em teoria, espécies com maior longevidade teriam uma seleção mais rígida contra retrovírus, durante a sua história evolutiva, refletindo em menor diversidade e número de sequências inseridas no genoma. Isso poderia explicar o baixo número de ERVs encontrados na onça, considerando-se que a amostragem obtida neste estudo não seria um artefato amostral. A confirmação dessa premissa poderia ser obtida se esse mesmo padrão fosse encontrado em outras espécies e em indivíduos dessa linhagem que ainda não foram estudados.

É muito difícil quantificar longevidade em espécies silvestres, entretanto, existe uma correlação entre a idade da maturidade sexual com a longevidade em várias espécies de mamíferos (Millar & Zammuto, 1983; Ricklefs, 2010). Em uma busca sobre idade da maturidade sexual de alguns gatos silvestres na literatura, encontramos as seguintes estimativas: *Leopardus tigrinus* (2 anos); *L. wiedii* (2 anos); *L. geoffroyi* (1,5 a 2 anos); *Puma yagouaroundi* (1,4 a 3 anos); *P. concolor* (1,5 a 3 anos); e *Panthera onca* (2 anos em fêmeas, 4 anos em machos) (Sunquist & Sunquist, 2002).

A maioria destas estimativas é baseada em relatos de alguns pesquisadores de campo e zoológicos. No entanto, é possível observar que *L. tigrinus*, *L. wiedii* e *L. geoffroyi* adquirem maturidade sexual em intervalos de idade similares. Portanto, longevidade não refletiria a diferença na diversidade e no número de ERVs encontrados nessas espécies. Embora esta observação seja de certa forma, especulativa devido ao

baixo número de espécies amostradas, este resultado seria o esperado, porque as inserções dos ERVs encontrados seriam muito antigas, refletindo marcas de infecções em seus ancestrais, como discutiremos mais adiante. Desse modo, a ecologia, os mecanismos de defesa (imunidade inata e adquirida) e a demografia desses ancestrais associados à dinâmica populacional das unidades (que hoje denominamos espécies) provavelmente influenciaram a diversidade dos ERVs encontrados neste estudo.

### **6.3. SINEs**

Um fato curioso constatado neste estudo foi identificar inserções de SINEs na região da transcriptase reversa dos clones LgEM10003 e PcoJO10006. Inserções de SINEs e retrotransposons em sequências de origem viral foram descritas em alguns estudos (Friesen & Nissen, 1990; Bénit *et al.* 2003; Piskurek & Okada 2007). Quando analisamos o SINE inserido na sequência LgEM10003, verificamos que ele não está presente nas sequências dos clones RV-cougar e RV-Domestic cat, cuja filogenia indica ancestralidade comum. Portanto, seria interessante avaliar se essa sequência está fixada na espécie, o que sugeriria que a inserção ocorreu depois da separação da linhagem da Jaguatirica (constituída por *Leopardus pardalis*, *L. wiedii*, *L. jacobita*, *L. colocolo*, *L. geoffroyi*, *L. tigrinus* e *L. gingna*) ou se é resultante de polimorfismo ou, ainda, corresponde a um evento isolado em uma população/indivíduo. O mesmo raciocínio pode ser feito para o SINE inserido na sequência de PcoJO10006, proveniente da espécie *Puma concolor*.

Não foi possível amplificar clone semelhante ao LgEM10003 nos indivíduos das demais espécies estudadas, provavelmente devido ao fato de este estudo ter sido planejado para obter uma amostragem das sequências de ERVs existentes nas espécies, não capturando toda a diversidade de retrovírus.

### **6.4. Identificação dos retrovírus endógenos e os desafios encontrados**

Embora o gene da polimerase seja amplamente utilizado devido ao grau de conservação, apenas um número limitado, de aproximadamente 175 resíduos de aminoácidos da região da transcriptase reversa, apresenta identidade em nível de gênero

(33% a 66%) (Vogt, 1997). Outros grandes obstáculos são as confusas e não unificadas classificação e nomenclatura dos retroelementos (Blomberg *et al.*, 2009).

A obtenção de algumas sequências curtas, e, conseqüentemente, a pouca informação contida nelas, dificultou uma classificação pontual de alguns clones em relação aos ERVs descritos na literatura. Além disso, algumas sequências de ERVs disponíveis no GenBank, bem como um certo número de sequências identificadas neste estudo (LwiJO7007, LgEM10003, PyEM13001, PyEM13003 e PyEM6002), são difíceis de classificar quanto ao gênero, pois o tempo de integração tende a apagar informações sobre sua ancestralidade. Diante do exposto, procurou-se, pelo menos, classificar as sequências quanto à classe e fazer inferências sobre diversidade e evolução desta, utilizando como ferramenta principal a filogenia.

O clone Lwi JO7007 ilustra claramente esse problema. Como foi demonstrado, o fragmento analisado era uma sequência curta que se agrupou com a classe I de ERVs junto com os *Gamma* e *Epsilonretrovirus*. Apesar da pouca informação contida no fragmento desta sequência, verificou-se que ela representa um retroelemento presente em vários mamíferos placentários, e, provavelmente, evoluiu de modo semelhante nesses animais.

## **6.5. Linhagem *Myotis***

Os resultados evidenciaram a existência de uma linhagem de retrovírus endógeno, amplamente difundida em várias espécies de felídeos.

A busca no banco de dados do NCBI, utilizando a sequência consenso como *query*, encontrou na espécie *Myotis lucifugus* a segunda mais alta identidade e número de resultados (hits) com a sequência consenso, apesar de WGS completas ou parciais estarem disponíveis para 19 ordens de mamíferos (<http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi>).

As filogenias também indicaram uma proximidade muito grande entre as sequências dos felídeos e *Myotis lucifugus*. Isso sugeriria transmissão interespecie de um retrovírus em um passado longínquo, possivelmente no ancestral das linhagens da Jaguaririca, Puma e Gato doméstico há aproximadamente oito milhões de anos, com base nas estimativas de Johnson *et al.* (2006).

*Myotis lucifugus* é uma espécie que habita a América do Norte e, da mesma forma que os felídeos americanos, ancestrais dessa espécie migraram da Eurásia para América, via estreito de Bering, na metade do Mioceno (Stadelmann *et al*, 2007). O fato de o gato doméstico possuir sequências desse retroelemento proximamente relacionadas com espécies de felídeos americanos leva a crer que a transmissão tenha ocorrido antes da migração do ancestral dessa espécie para a América, quando habitavam a Eurásia. Este cenário poderia ser confirmado através da ausência de sequências similares nas linhagens Caracal e Bay cat. O padrão monofilético das sequências observado na filogenia com retenção de linhagens ancestrais (LPJO9 e Lwi JO7005) indica que as sequências desse retroelemento estariam (co) evoluindo com os felídeos há algum tempo, mas não o suficiente para apagar informações de inserções iniciais. Apesar de o agrupamento das sequências LPJO9 e Lwi JO7005 apresentar baixo suporte de *bootstrap*, *L. pardalis* e *L. wiedii* são espécies evolutivamente muito próximas, consideradas espécies-irmãs (Jonhson *et al* 2006). Portanto, o fato de possuírem sequências similares é possivelmente um reflexo de suas trajetórias evolutivas.

## 6.6. Relações filogenéticas

As análises apresentadas neste estudo fornecem exemplo de um gênero de retrovírus (*Gammaretrovirus*) que infectou um grupo de espécies filogeneticamente muito próximas, ao longo de suas histórias evolutivas, em acordo com vários estudos descritos em vertebrados (Herniou *et al.*, 1998; Bénit *et al.*, 2003; Jaratlerdsiri *et al.*, 2009). Os *Gammaretrovirus* estão amplamente difundidos no genoma das espécies estudadas, mas a sua real diversidade ao longo da história provavelmente está subestimada. Em primeiro lugar, foram estudados apenas registros fósseis, pois aqueles vírus, que não foram integrados em uma célula germinativa do hospedeiro, não estão representados. Em segundo lugar, foi sequenciada uma amostragem de clones por espécie, não representando a totalidade da diversidade contida em seus genomas. Apesar disso, foram identificados três grupos de *Gammaretrovirus* evoluindo de modo particular nesse grupo de espécies.

O grupo G3 é aparentemente o mais recente, apresentando o menor número de códons de parada e/ou *indels* e menor distância genética intragrupo. A família ERV-Fc de retrovírus originários de primatas, que está proximamente relacionada aos retrovírus

desse grupo de felídeos, apresenta baixo número de cópias e, aparentemente, uma recente endogenização caracterizada pela presença de provírus com ORFs intactas (Bénit *et al.*, 2003), semelhante ao que foi sugerido para o grupo G3.

Diferentemente do grupo G3, o grupo G2 reflete uma integração antiga, apresentando todas as sequências defectivas, de modo que estaria aparentemente extinto. O seu clado mais especioso está estreitamente relacionado com outros carnívoros que provavelmente compartilham receptores similares. Portanto, aparenta ser um clado mais especializado em termos de hospedeiros. Observa-se, na filogenia (Figura 20), que o grupo sofreu uma expansão quase simultânea, característica de organismos que ocupam novos nichos ecológicos, o que pode ser representado no mundo molecular pela disponibilidade de receptores que podem estar sendo utilizados, na atualidade, por outros retroelementos. Próximo a este clado, na filogenia da figura 21, encontra-se um clado de primatas com características semelhantes em relação à especificidade dos hospedeiros.

As sequências amostradas do grupo G1 apresentam características intermediárias em relação ao número de restrições funcionais (Figura 20B2), sendo que a linhagem representada pela sequência LgJO10006 ainda poderia estar ativa em alguma espécie. A grande maioria das linhagens do grupo G1 é compartilhada com outros mamíferos, sugerindo ser mais generalista em termos de hospedeiros. Nas figuras 20 e 21, observa-se que, proximamente relacionadas a essas sequências, encontram-se os retrovírus FeLV e KoEV que são muito patogênicos para felídeos e coalas, respectivamente.

A filogenia molecular representada pela Figura 21 sugere a ocorrência de nove integrações independentes de ERVs *Gamma-like* no genoma de felídeos. Sete delas compartilham sequências entre *F. catus* e espécies neotropicais. A distância genética observada dentro dos sete clados de ERVs entre *F. catus* e as demais espécies de felídeos neotropicais variou de 0,028 a 0,317. Esta distância, no seu limite inferior, é similar ou maior do que distâncias observadas entre as espécies neotropicais e *F. catus* quando foram analisados fragmentos dos genes nucleares CHRNA1, FES e GHR. No seu limite superior, a distância apresenta-se maior que as observadas com o gene mitocondrial *Citocromo b*, que apresenta uma taxa evolutiva mais alta que a dos genes nucleares. Estes valores de distâncias genéticas elevados, em comparação com outros genes, sugerem integrações antigas. Além disso, o gato doméstico foi introduzido na

América, durante a colonização, por imigrantes europeus (Shostell *et al.* 2005) e portanto, o contato entre felídeos neotropicais e gatos domésticos teria acontecido muito recentemente, em uma escala de apenas centenas de anos. Logo, as sete distintas integrações teriam ocorrido em espécies ancestrais, e as atuais espécies de felídeos estariam compartilhando os registros fósseis destas integrações.

As outras três linhagens de ERVs distintos de *Gammaretrovirus* presentes no genoma do *L. geoffroyi*, *P. yagouaroundi* e *L. wiedii* indicam que retrovírus similares a outros gêneros também infectavam ancestrais destas espécies em um passado longínquo.

### **6.7. Recombinação**

Os retrovírus possuem a reputação de evoluírem sob a influência de recombinação. Observando o ciclo de vida de um retrovírus, é difícil imaginar que esses organismos não troquem material genético. Entretanto, simplesmente trocar material genético não significa sucesso em um evento de recombinação (Worobey & Holmes 1999).

O fato de não ter sido observado um maior número de eventos recombinantes não significa que esses eventos não ocorreram durante a evolução destes retroelementos. O mais correto seria pensar que falhamos em rejeitar a hipótese nula que prediz a ausência de recombinação. Em primeiro lugar, clones recombinantes dos grupos G1 e G2 podem não ter sido amostrados. Em segundo lugar, esses clones podem ter sido extintos devido a eventos estocásticos ou de seleção purificadora.

### **6.8. Evolução de ERVs através de uma visão contemporânea**

Emerman & Malik (2010) sugeriram que a seleção para sobreviver aos efeitos patogênicos dos *paleovirus* (vírus ancestrais) moldou o nosso repertório de defesas antivirais, influenciando a nossa resistência e susceptibilidade aos vírus contemporâneos. Este estudo exemplificou infecções de retrovírus ancestrais, demonstrando que alguns arranjos representados na filogenia da Figura 21 poderiam ser atribuídos à transmissão de vírus similares entre diferentes ordens de mamíferos, por exemplo, Carnivora e Chiroptera. Diante do exposto, ressalta-se o fato, hoje

amplamente aceito pela ciência, que deveríamos avaliar riscos de surgimento de vírus emergentes que poderiam transpor barreiras entre hospedeiros de espécies distintas, quando “forçamos” a convivência de espécies que não compartilham um longo período (histórico evolutivo) de coexistência. Este estudo também demonstrou ERVs evoluindo de modo idiossincrático como, por exemplo: ERVs com inserção de sequências de SINEs, o padrão evolutivo do grupo *Myotis* e as peculiaridades dentro dos três principais grupos de ERVs *Gamma-like*. Portanto, outro aspecto importante a considerar é que, através do uso de alguns tipos de vacina produzidas a partir de linhagens celulares que produzem partículas virais de ERVs, poderíamos estar introduzindo ERVs que não fazem parte da história evolutiva de uma determinada categoria taxonômica. Procedendo desta forma, estaríamos agindo como se estivéssemos misturando objetos de sítios arqueológicos diferentes. Deste modo, as amostras de zoológicos ou animais de cativeiro utilizadas em estudos de TEs deveriam ser avaliadas com relação a este critério.

## 7. CONCLUSÃO

Neste estudo foram encontrados ERVs em todas as espécies de felídeos estudadas, semelhantes ao descrito em outros vertebrados. O padrão de retroelementos similares a retrovírus encontrados reflete que estes organismos têm acompanhado essas espécies de felídeos durante a sua história evolutiva. Na atualidade, aparentemente, os retrovírus endógenos são registros fósseis de infecções ocorridas no passado, mas também poderiam ser considerados como um estoque de diversidade que, em teoria, poderia ser utilizado por um retroelemento ativo através de eventos de recombinação. Portanto, a sua história evolutiva não pode ser vista simplesmente como um réquiem.

Observou-se dominância de ERVs similares a *Gammaretrovirus*, bem como semelhanças na composição desses retroelementos em alguns mamíferos. Os ERVs similares a *Gammaretrovirus* amostrados neste estudo foram divididos em três grandes grupos e, nesses grupos, os ERVs pertencentes às diferentes espécies apresentam-se distribuídos de forma heterogênea, dificultando a observação de um padrão coevolutivo. Entretanto, numa visão mais pontual nas filogenias, observou-se que estes retroelementos tendiam, em alguns momentos, a acompanhar a evolução de seus hospedeiros.

Foi possível, também, observar algumas sequências isoladas que se agruparam próximo a *Betaretrovirus* e a *Epsilonretrovirus*. Isso indica que ancestrais desses felídeos também teriam sido infectados por outros gêneros, além do *Gammaretrovirus*.



## 8. REFERÊNCIAS BIBLIOGRÁFICAS

ABASCAL, F.; ZARDOYA R.; POSADA, D. ProtTest: selection of best-fit models of protein evolution. **Bioinformatics**. v. 21, p. 2104-2105, 2005.

AHMAD, S.; LEVY, L. S. The Frequency of Occurrence and Nature of Recombinant Feline Leukemia Viruses in the Induction of Multicentric Lymphoma by Infection of the Domestic Cat with FeLV-945. **Virology**. v. 403, n. 2, p.103-110, 2010.

ALTSCHUL, S. F.; MADDEN, T. L.; SCHÄFFER, A. A.; ZHANG, J.; ZHANG, Z.; MILLER, W; LIPMAN, D. J. "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", **Nucleic Acids Research**. v. 25, p. 3389-3402, 1997.

ANISIMOVA, M.; NIELSEN, R.; YANG, Z. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. **Genetics**. v.164, p. 1229-1236, 2003.

ARNAUD, F. et al. A Paradigm for Virus–Host Coevolution: Sequential Counter-Adaptations between Endogenous and Exogenous Retroviruses. **Plos Pathogens**. v. 3, n. 11, p. 1716-1729, 2007.

BANNERT, N.; FIEBIG, U.; HOHN, O. Retroviral particles, proteins and genomes. In: Kurth R.; Bannert, N. (Eds). **Retroviruses: Molecular Biology, Genomics and Pathogenesis**. Caister Academic Press. p. 71-106, 2010.

BELSHAW, R.; PEREIRA, V.; KATZOURAKIS, A.; TALBOT, G.; PACES, J.; BURT, A.; TRISTEM, M. Long-term reinfection of the human genome by endogenous retroviruses. **Proceedings of the National Academy of Sciences of the United States of America**. v.101, p. 4894-4899, 2004.

BÉNIT, L.; CALTEAU, A.; HEIDMANN, T. Characterization of the low-copy HERV-Fc family: evidence for recent integrations in primates of elements with coding envelope genes. **Virology**. v. 312, p.159-168, 2003.

BIEK, R.; RODRIGO, A. G.; HOLLEY, D.; DRUMMOND, A.; ANDERSON, C. R. J. R.; ROSS, H. A.; POSS, M. Epidemiology, Genetic Diversity, and Evolution of Endemic Feline Immunodeficiency Virus in a Population of Wild Cougars. **Journal of Virology**. v. 77, p. 9578-9589, 2003.

BLIKSTAD, V.; BENACHENHOU, F.; SPERBER, G.O.; BLOMBERG, J. Evolution of human endogenous retroviral sequences: a conceptual account. **Cellular and Molecular Life Sciences**. v. 65, p. 3348-3365, 2008.

BLOMBERG, J.; BENACHENHOU, F.; BLIKSTAD, V.; SPERBER, G; MAYER, J. Classification and nomenclature of endogenous retroviral sequences (ERVs): problems and recommendations. **Gene**. v. 448, n. 2, p. 115-123, 2009.

BLOMBERG, J.; SHEIKHOLVAEZIN, A.; ELFAITOURI, A.; BLOMBERG, F.; SJÖSTEN, A.; ULFSTEDT, J.M.; PIPKORN, R.; KÄLLANDER, C.; ÖHRMÄLM, C., SPERBER, G. Phylogeny-directed search for murine leukemia virus - like retroviruses in vertebrate genomes, and in patients suffering from Myalgic encephalomyelitis/Chronic Fatigue Syndrome and Prostate cancer. **Advances in Virology**. v. 2011, Article ID 341294, 2011.

BOTTU, G. Sequence databases and database searching. In: **The Phylogenetic Handbook**. Lemey, P.; Salemi, M.; Vandamme A. M. (eds). Cambridge University Press. p. 33-54, 2009.

BRUEN, T.; PHILIPPE, H.; BRYANT, D. A simple and robust statistical test for detecting the presence of recombination. **Genetics**. v.172, p. 2665–2681, 2006.

CHESSA, B. et al. Revealing the History of Sheep Domestication Using Retrovirus Integrations. **Science**. v. 324, p. 532-536, 2009.

DENNER, J. Endogenous Retrovirus. In: **Retroviruses: Molecular biology, genomics and Pathogenesis**. Kurth, R.; Bannert, N. (eds). Caister Academic Press. p. 35-69, 2010.

DEWANNIEUX, M.; RIBET, D.; HEIDMANN, T. “Risks linked to endogenous retroviruses for vaccine production: a general overview” **Biologicals**. v. 38, n. 3, p. 366-370, 2010.

DO, C. B.; MAHABHASHYAM, M. S., BRUDNO, M.; BATZOGLOU, S. ProbCons: Probabilistic consistency-based multiple sequence alignment. **Genome Research**. v.15, p. 330-340, 2005.

DOXIADIS, G. G.; DE GROOT, N.; BONTROP, R. E. Impact of endogenous intronic retroviruses on major histocompatibility complex class II diversity and stability. **Journal of Virology**. v. 82, n. 13, p. 6667-77, 2008.

EMERMAN, M.; MALIK, H. S. Paleovirology-modern consequences of ancient viruses. **PLoS Biology**. v. 8, n. 2, 2010.

FRANKLIN, S. P.; TROYER J. L.; TERWEE J. A.; LYREN, L. M.; BOYCE, W. M.; RILEY, S. P. D.; ROELKE, M. E.; CROOKS, K. R.; VANDEWOUDE, S. Frequent Transmission of Immunodeficiency Viruses among Bobcats and Pumas. **Journal of Virology**. v. 81, n. 20, p. 10961-10969, 2007.

FRIESEN, P. D.; NISSEN, M. S. Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome. **Molecular and Cellular Biology**. v.10 , n. 6, p. 3067-3077, 1990.

GIFFORD, R.; TRISTEM, M. The evolution, distribution and diversity of endogenous retroviruses. **Virus Genes**. v. 26, n. 3, p. 291-315, 2003.

GIFFORD, R. J.; KABAT, P.; MARTIN, J.; LYNCH, C.; TRISTEM, M. Evolution and Distribution of Class II-Related Endogenous Retroviruses. **Journal of Virology**. v. 79, n. 10, p. 6478-6486, 2005.

GIFFORD, R. J.; KATZOURAKIS, A.; TRISTEM, M.; PYBUS, O. G.; WINTERS, M.; SHAFER, R. W. A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. **Proceedings of the National Academy of Sciences of the United States of America**. v. 105, p. 20362-20367, 2008.

GOFF, S. P. Retroviridae: The Retroviruses and Their Replication. In: **Fields Virology**. Knipe D. M.; Howley P. M. (eds.). p.1871-1940. Lippincott, 2001.

GUINDON S.; GASCUEL O. PhyML: "A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood". **Systematic Biology**. v. 52, n. 5, p. 696-704, 2003.

HALL, T. A. BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symposium Series**. v. 41, p. 95-98, 1999.

HARRIS, R. S.; LIDDAMENT, M. T. Retroviral restriction by APOBEC proteins. **Nature Reviews Immunology**. v. 4, p. 868-877, 2004

HERNIOU, E.; MARTIN, J.; MILLER, K.; COOK, J.; WILKINSON, M.; TRISTEM, M. Retroviral diversity and distribution in vertebrates. **Journal of Virology**. v. 72, p. 5955-5966, 1998.

HOLMES, E. C. The phylogeography of human viroses. **Molecular Ecology**. v. 13, p. 745-756, 2004.

HOLMES, E. C. The comparative genomics of viral emergence. **Proceedings of the National Academy of Sciences of the United States of America**. v. 107, p. 1742-1746, 2010.

HOLMES, E. C.; WOROBEY, M.; RAMBAUT, A. Phylogenetic evidence for recombination in dengue virus. **Molecular Biology and Evolution**. v. 16, p. 405-409, 1999.

HUÉ, S. et al. Disease-associated XMRV sequences are consistent with laboratory contamination. **Retrovirology**. v. 7, n. 111, 2010.

HUELSENBECK, J. P.; RONQUIST, F. MrBayes: Bayesian inference of phylogeny. **Bioinformatics**. v.17, p. 754-755, 2001.

HUSON, D. H.; BRYANT, D. Application of Phylogenetic Networks in Evolutionary Studies. **Molecular Biology and Evolution**. v. 23, n. 2, p. 254-267, 2006.

HUYSE, T.; POULIN, R.; THÉRON, A. Speciation in parasites: a population genetics approach. **Trends in Parasitology**. v. 21, p. 469-475, 2005.

INDRUSIAK, C.; EIZIRIK, E. Carnívoros. In: **Livro vermelho da fauna ameaçada de extinção no Rio Grande do Sul**. Fontana, C. S.; Benck, G. A.; Reis, R. E. (eds). EDIPUCRS. p. 507-533, 2003.

JARATLERDSIRI, W.; RODRIGUEZ-ZÁRATE, C. J.; ISBERG S. R.; DAMAYANTI, C. S.; MILES, L. G.; CHANSUE, N.; MORAN, C.; MELVILLE, L.; GONGORA, J. Distribution of Endogenous Retroviruses in Crocodylians. **Journal of Virology**. v. 83, n. 19, p.10305-10308, 2009.

JERN, P.; COFFIN, J. M. Effects of retroviruses on host genome function. **Annual Review of Genetics**. v. 42, p. 709-732, 2008.

JERN, P.; SPERBER, G. O.; BLOMBERG, J. Use of endogenous retroviral sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. **Retrovirology**. v. 2, n. 50, 2005.

JOHNSON, .W. E.; EIZIRIK, E.; PECON-SLATTERY, J.; MURPHY, W.J.; ANTUNES, A.; TEELING, E.; O'BRIEN, S. J. The Late Miocene Radiation of Modern Felidae: A Genetic Assessment. **Science**. v. 311, p. 73-77, 2006.

KATZOURAKIS, A.; TRISTEM, M.; PYBUS, O. G.; GIFFORD, R. J. Discovery and analysis of the first endogenous lentivirus. **Proceedings of the National Academy of Sciences of the United States of America**. v. 104, p. 6261-6265, 2007.

KAZAZIAN, H. H. Jr. Mobile elements: Drivers of genome evolution. **Science**. v. 303, p. 1626-1632, 2004.

KECKESOVA, Z.; YLINEN, L. M. J.; TOWERS, G. J.; GIFFORD R. J.; KATZOURAKIS, A. Identification of a RELIK orthologue in the European hare (*Lepus europaeus*) reveals a minimum age of 12 million years for the lagomorph lentiviruses. **Virology**. v. 384, p. 7-11, 2009.

LEMEY, P.; POSADA, D. Introduction to recombination detection. In: **The Phylogenetic Handbook**. Lemey, P.; Salemi, M.; Vandamme A. M. (eds.). Cambridge University Press. p. 493-518, 2009.

LEWIN, B. **Genes IX**. Jones and Bartlet Publishers, 892p., 2008.

LIBRADO, P.; ROZAS, J. DnaSP v5: A software for comprehensive analysis of Dna polymorphism data. **Bioinformatics**. v. 25, p. 1451-1452, 2009.

LOISEL, D. A.; ALBERTS, S. C.; OBER, C. Functional significance of MHC variation in mate choice, reproductive outcome, and disease risk. In: **Evolution in health and disease**. Stearns, S.; Koella, J. C. P. (eds.). Oxford University Press. p. 95-108, 2008.

LÖWER, R.; LÖWER, J.; KURTH, R. The viruses in all of us: Characteristics and biological significance of human endogenous retrovirus sequences. **Proceedings of the National Academy of Sciences of the United States of America**. v. 93, p. 5177-5184, 1996.

MAKSAKOVA, I. A.; MAGER, D. L.; REISS, D. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. **Cellular and Molecular Life Sciences**.v. 65, p. 3329-3347, 2008.

MATTERN, M. Y.; MCLENNAN, D. A. Phylogeny and Speciation of Felids. **Cladistics**. v. 16, p. 232-253. 2000.

MAYER, J.; BLOMBERG, J.; SEAL, R. L. A revised nomenclature for transcribed human endogenous retrovirus loci. **Mobile DNA**. v. 2, n. 7, 2011.

McCLINTOCK, B. The origin and behavior of mutable loci in maize. **Proceedings of the National Academy of Sciences of the United States of America**. v. 36, p. 344-355. 1950.

MILLAR, J. S.; ZAMMUTO, R. M. Life history of mammals: an analysis os life tables. **Ecology**. v. 64, p. 631-635, 1983.

MIYAZAWA, T. Endogenous retroviruses as potential hazards for vaccines. **Biologicals**. v. 38, p. 371-376, 2010.

MIYAZAWA, T.; YOSHIKAWA, R.; GOLDBER, M.; OKADA, M.; STEWART, H.; PALMARINI, M. Isolation of an infectious endogenous retrovirus in a proportion of live attenuated vaccines for pets. **Journal of Virology**. v. 84, p. 3690-3694, 2010.

MORGENSTERN, B. DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. **Bioinformatics**. v.15, p. 211- 218, 1999.

MULLER, H. J. The relation of recombination to mutational advance. **Mutation Research**. v.1, p. 2-9, 1964.

MÜNK, C.; BECK T.; ZIELONKA, J.; HOTZ-WAGENBLATT, A.; CHAREZA, S.; BATTENBERG, M.; THIELEBEIN, J.; CICHUTEK, K.; BRAVO, I. G.; O'BRIEN, S. J.; LOCHELT, M.; YUHKI, N. Functions, structure, and read-through alternative splicing of feline APOBEC3 genes. **Genome Biology**. v. 9 (R48), 2008.

NARUSHIMA, R.; HORIUCHI, N.; USUI, T.; OGAWA, T.; TAKAHASHI, T.; SHIMAZAKI, T. Experimental infection of dogs with a feline endogenous retrovirus RD-114. **Acta Veterinaria Scandinavica**. v. 53, n. 3, 2011.

O'BRIEN, S. J.; TROYER, J.; ROELKE, M.; MARKER, L.; PECON-SLATTERY, J. Plagues and adaptations: Lessons from the Felidae modes for AIDS and SARS. **Biological Conservation**. v. 131, p. 255-267, 2006.

OVERBAUGH, J.; BANGHAM, C. R. M. Selection Forces and Constraints on Retroviral Sequence Variation. **Science**. v. 292, p. 1106-1109, 2001.

PAGE, R. D. M., HOLMES, E. C. **Molecular evolution: a phylogenetic approach**. Blackwell Science Ltd., 1998.

- PECON-SLATTERY, J.; MURPHY, W. J.; O'BRIEN, S. J. Patterns of diversity among SINE elements isolated from three Y-chromosome genes in carnivores. **Molecular Biology and Evolution**. v. 17, p. 825-829, 2000.
- PERTSEMLIDIS, A.; FONDON, J. W. III. Having a BLAST with bioinformatics (and avoiding BLASTphemy). **Genome Biology**. v. 2, reviews 2002.1–reviews 2002.10, 2001.
- POLANI, S.; ROCA, A. L.; ROSENSTEEL, B. B.; KOLOKOTRONIS, S. O.; BARGAL, G. K. Evolutionary dynamics of endogenous feline leukemia virus proliferation among species of the domestic cat lineage. **Virology**. 405, p. 397-407, 2010.
- POND, S. L. K.; FROST, S. D. W.; MUSE, S. V. HyPhy: hypothesis testing using phylogenies. **Bioinformatics**. v. 21, n. 5, p. 676-679, 2005.
- PONTIUS, J. U. et al. Initial sequence and comparative analysis of the cat genome. **Genome Research**. v. 17, p. 1675-1689, 2007.
- POSADA, D. jModelTest: Phylogenetic Model Averaging. **Molecular Biology and Evolution**. v. 25, p.1253-1256, 2008.
- POSS, M.; BIEK, R.; RODRIGO, A. Viruses as Evolutionary Tool to Monitor Populations Dynamic. In: **Conservation Medicine: Ecological Health in Practice**. Aguirre, A. A.; Ostfeld, R. S.; Tabor, G. M.; House, C.; Pearl, C. M. (eds) .Oxford University Press. p 118-144, 2002.
- POSS, M.; IDOINE, A.; ROSS, H. A.; TERWEE, J. A.; VANDEWOUDE, S.; RODRIGO, A. Recombination in feline lentiviral genomes during experimental cross-species infection. **Virology**. v. 359, p. 146-151, 2007.
- PISKUREK, O.; OKADA, N. Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. **Proceedings of the National Academy of Sciences of the United States of America**. v. 104, p. 12046-12051, 2007.
- PYBUS, O. G.; SHAPIRO, B. Natural selection and adaptation of molecular sequences. In: **The Phylogenetic Handbook**. Lemey, P.; Salemi, M.; Vandamme A. M. (eds). Cambridge University Press. p. 407-418, 2009.
- RAVAZZOLO, A. P.; COSTA, U. M. Virologia Especial- *Retroviridae*. In: **Virologia Veterinária**. Flores, E. F. (Org.). Editora da Universidade Federal de Santa Maria, p. 809-837, 2007.
- RAY, D. A.; BATZER, M. A. Reading TE leaves: New approaches to the identification of transposable element insertions. **Genome Research**. v. 21, p. 813-820, 2011.
- RICKLEFS, R. E. Life-history connections to rates of aging in terrestrial vertebrates. **Proceedings of the National Academy of Sciences of the United States of America**. v. 107, n. 22, p. 10314-10319, 2010.

ROMANO, C. M.; DE MELO, F. L.; CORSINI, M. A. B.; HOLMES, E. C.; ZANOTTO, P. M. Demographic Histories of ERV-K in Humans, Chimpanzees and Rhesus Monkeys. **PLoS ONE**. v. 2, n. 10, 2007.

SAMBROOK, J.; FRITSCH, E. F.; MANIATS, T. **Molecular cloning: a laboratory manual**. Cold Spring Harbour Laboratory. v. 3, p. 37, 1989.

SCHAACK, S.; GILBERT, C.; FESCHOTTE, C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. **Trends in Ecology and Evolution**. v. 25, n. 9, p. 537-546, 2010.

SINZELLE, L.; CARRADE, C. Q.; PAILLARD, E.; BRONCHAIN, O. J.; POLLET, N. Characterization of a *Xenopus tropicalis* Endogenous Retrovirus with Developmental and Stress-Dependent Expression. **Journal of Virology**. v. 85, n. 5, p. 2167-2179, 2011.

STADELMANN, B.; LIN, L. K.; KUNZ, T. H., RUEDI, M. Molecular phylogeny of New World *Myotis* (Chiroptera, Vespertilionidae) inferred from mitochondrial and nuclear DNA genes. **Molecular Phylogenetics and Evolution**. v. 43, p. 32-48, 2007.

SULLIVAN, J.; JOYCE, P. Model selection in phylogenetics. **Annual Review of Ecology, Evolution, and Systematics**. v. 36, p. 445-466, 2005.

SUNQUIST, M.; SUNQUIST, F. **Wild Cats of the World**. University of Chicago Press, 452 p., 2002.

SYMER, D. E.; BOEKE, J. D. An everlasting war dance between retrotransposons and their metazoan host. In: **Retroviruses: Molecular biology, genomics and Pathogenesis**. Kurth, R.; Bannert, N. (eds). Caister Academic Press. p. 1-33, 2010.

TAJIMA, F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. **Genetics**. v. 123, p. 585-595, 1989.

TAMURA, K.; PETERSON, D.; PETERSON, N.; STECHER, G.; NEI, M.; KUMAR, S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. **Molecular Biology and Evolution**. v. 28, n. 10, p. 2731-9, 2011.

TANDON, R.; CATTORI, V.; WILLI, B.; LUTZ, H.; HOFMANN-LEHMANN, R. Quantification of endogenous and exogenous feline leukemia virus sequences by real-time PCR assays. **Veterinary Immunology and Immunopathology**. v. 123. p. 129-133, 2008.

TARLINTON, R. E.; MEERS, J.; YOUNG, P. R. Retroviral invasion of the koala genome. **Nature**. v. 442, n. 6, p. 79-81, 2006.

TRIGO, T. C.; FREITAS, T. R. O.; KUNZLER, G.; CARDOSO, L.; SILVA, J. C. R., JOHNSON, W. E.; O'BRIEN, S. J.; BONATTO, S. L.; EIZIRIK, E. Inter-species hybridization among Neotropical cats of the genus *Leopardus*, and evidence for an introgressive hybrid zone between *L. geoffroyi* and *L. tigrinus* in southern Brazil. **Molecular Ecology**. v. 17, p. 4317-4333, 2008.

TRISTEM, M. Amplification of divergent retroelements by PCR. **BioTechniques**. v. 20, p. 608-612, 1996.

TRISTEM, M.; KABAT, P.; LIEBERMAN, L.; LINDE, S.; KARPAS, A.; HILL, F. Characterisation of a novel murine leukemia virus-related subgroup within mammals. **Journal of Virology**. v.70, p. 8241-8246, 1996.

VAN DER KUYL, A. C.; DEKKER, J. T.; GOUDSMIT, J. Distribution of baboon endogenous virus among species of African monkeys suggests multiple ancient cross-species transmissions in shared habitats. **Journal of Virology**. v. 69, p. 7877-7887, 1995.

WEISS, R. A. The discovery of endogenous retroviruses. **Retrovirology**. v. 3, p. 67, 2006.

WICKER, T. et al. A unified classification system for eukaryotic transposable elements. **Nature Review Genetic**. v. 8, p. 973-982, 2007.

WOOLHOUSE, M. E. J.; HAYDON, D. T.; RUSTOM, A. Emerging pathogens: the epidemiology of species jump. **Trends in Ecology and Evolution**. v. 20, p. 238-244, 2005.

WOROBAY, M.; HOLMES, E. C. Evolutionary aspects of recombination in RNA viruses. **Journal of General Virology**. v. 80, p. 2535-2543, 1999.

YUHKI, N.; MULLIKIN, J. C.; BECK, T.; STEPHENS, R.; O'BRIEN, S. J. Sequences, Annotation and Single Nucleotide Polymorphism of the Major Histocompatibility Complex in the Domestic Cat. **PLoS ONE**. v. 3, n. 7, 2008.

XIA, X. Assessing substitution saturation with DAMBE. In: **The Phylogenetic Handbook**. Lemey, P.; Salemi, M.; Vandamme, A. M. (eds.). Cambridge University Press. p. 615-630, 2009.

XIA, X.; XIE, Z. DAMBE: data analysis in molecular biology and evolution. **Journal of Heredity**. v. 92, p. 371-373, 2001.

XIAO, R.; PARK, K.; LEE, H.; KIM, J.; PARK, C. Identification and Classification of Endogenous Retroviruses in Cattle. **Journal of Virology**. v. 82, p. 582-587, 2008.

XIONG, Y.; EICKBUSH, T. H. Origin and evolution of retroelements based upon their reverse transcriptase. **The EMBO Journal**. v. 9, p. 3353-3362, 1990.



## 9. APÊNDICES

**Apêndice A.** Lista das sequências obtidas no GenBank e utilizadas nas análises filogenéticas, seus respectivos números de identificação, a sigla utilizada nas filogenias, classificação e hospedeiro.

<b>Retrovírus</b>	<b>ID (GenBank)</b>	<b>Sigla</b>	<b>Classificação</b>	<b>Classe ERV</b>	<b>Hospedeiro</b>
Avian leukosis virus	AAQ55056.1	ALV	<i>Alpharetrovirus</i>		<i>Gallus gallus</i>
Bovine leukemia virus	D00647.1	BLV	<i>Deltaretrovirus</i>		<i>Bos taurus</i>
Canis lupus familiaris- DERV3	HM460340.1	DERV3	<i>Gamma-like</i>	I	<i>Canis lupus familiaris</i>
Caprine arthritis encephalitis virus	M33677.1	CAEV	<i>Lentivirus</i>		<i>Capra hircus</i>
Casuarius casuarius endogenous virus	AY820067.1	Casuarius	-	II	<i>Casuarius casuarius</i>
Cat ECE1 endogenous retroviral	X51929.1	Cat ECE1	<i>Gamma-like</i>	I	<i>Felis silvestris</i>
Chimpanzee endogenous retrovirus	AJ507127.1	ERV-Fc1	-		<i>Pan troglodytes</i>
Endogenous retrovirus ZFERV	AF503912.1	ZFERV	<i>Epsilon-like</i>	I	<i>Danio rerio</i>
Feline foamy virus	U78765.1	FeSFV	<i>Spumavirus</i>		<i>Felis catus</i>
Feline immunodeficiency virus	M25381.1	FIV	<i>Lentivirus</i>		<i>Felis catus</i>
Feline leukemia virus	M18247.1	FeLV	<i>Gammaretrovirus</i>		<i>Felis catus</i>
<i>Felis catus</i> endogenous virus class II	AY820046.1	RV-Domestic cat	-	II	<i>Felis catus</i>
FERVmlu1 - <i>Felis catus</i> endogenous virus	EU153401.1	FERVmlu1	<i>Gamma-like</i>	I	<i>Felis catus</i>

Apêndice A. Continuação

<b>Retrovírus</b>	<b>ID (GenBank)</b>	<b>Sigla</b>	<b>Classificação</b>	<b>Classe ERV</b>	<b>Hospedeiro</b>
FERVmlu2 - <i>Felis catus</i> endogenous virus	EU153401.1	FERVmlu2	<i>Gamma-like</i>	I	<i>Felis catus</i>
<i>Gorilla</i> endogenous retrovirus	AJ507117.1	ERV-Fc2	<i>Gamma-like</i>	I	<i>Gorilla gorilla</i>
HERVK22	AC068700.5	HERVK22	<i>Betaretrovirus</i>	II	<i>Homo sapiens</i>
<i>Homo sapiens</i> - locus HUMERGPE	AAA73090.1	HUMERGPE	<i>Gamma-like</i>	I	<i>Homo sapiens</i>
<i>Homo sapiens</i> , retrovirus-like	D11078.1	HERV-H	<i>Gamma-like</i>	I	<i>Homo sapiens</i>
<i>Human endogenous retrovirus</i> HCML-ARV	AAP06677.1	HCML-ARV	<i>Gamma-like</i>	I	<i>Homo sapiens</i>
<i>Human T-lymphotropic virus 1</i>	ABM66584.1	HTLV-1	<i>Deltaretrovirus</i>		<i>Homo sapiens</i>
<i>Jaagsiekte sheep retrovirus</i>	EF680312.1	enJSRV	<i>Betarerovirus</i>	II	<i>Ovis aries</i>
<i>Mouse mammary tumor virus</i>	NP955564.1	MMTV	<i>Betaretrovirus</i>	II	<i>Mus musculus</i>
<i>Multiple Sclerosis Associated Retrovirus</i>	AAB66528.1	MSRV	<i>Gamma-like</i>	I	<i>Homo sapiens</i>
<i>Murine endogenous retrovirus ERV-L</i>	CAA73251.1	MuERV-L	-	III	<i>Mus musculus</i>
<i>Murine leukemia related virus (strain BoEV)</i>	X99924.1	BoEV	<i>Gamma-like</i>	I	<i>Bos taurus</i>
<i>Murine leukemia related virus (strain MeEVII)</i>	X99927.1	MeEVII	<i>Gamma-like</i>	I	<i>Meles meles</i>
<i>Murine leukemia related virus (strain MeEVI)</i>	X99926.1	MeEVI	<i>Gamma-like</i>	I	<i>Meles meles</i>
<i>Murine leukemia related virus (strain MiEVI)</i>	X99931.1	MiEVI	<i>Gamma-like</i>	I	<i>Ovis aries</i>
<i>Murine leukemia related virus (strain MiEVII)</i>	X99929.1	MiEVII	<i>Gamma-like</i>	I	<i>Mustela vison</i>

Apêndice A. Continuação

<b>Retrovírus</b>	<b>ID (GenBank)</b>	<b>Sigla</b>	<b>Classificação</b>	<b>Classe ERV</b>	<b>Hospedeiro</b>
<i>Murine leukemia related virus (strain OvEVII)</i>	X99932.1	OvEVII	<i>Gamma-like</i>	I	<i>Ovis aries</i>
<i>Murine leukemia related virus (strain TaEV)</i>	X99934.1	TaEV	<i>Gamma-like</i>	I	<i>Tadarida brasiliensis</i>
<i>Murine leukemia related virus (strain VuEV)</i>	X99935.1	VuEV	<i>Gamma-like</i>	I	<i>Vulpes vulpes</i>
<i>Murine leukemia virus (strain HaEV)</i>	X99925.1	HaEV	<i>Gamma-like</i>	I	<i>Halichoerus grypus</i>
<i>Mus dunni endogenous virus</i>	AAC31805.1	MDEV	<i>Gamma-like</i>	I	<i>Mus dunni</i>
<i>Nyctomys sumichrasti endogenous virus</i>	HM160041.1	Nyctomys	-	-	<i>Nyctomys sumichrasti</i>
<i>Oryzomys palustris endogenous virus</i>	DQ139767.1	Oryzomys	-	-	<i>Oryzomys palustris</i>
<i>Phascolarctos cinereus retrovirus</i>	AF151794.2	KoEV	<i>Gamma-like</i>	I	<i>Phascolarctos cinereus</i>
<i>Puma concolor endogenous virus class II</i>	AY820105.1	RV-Cougar	-	II	<i>Puma concolor</i>
<i>Python curtus endogenous retrovirus</i>	AF500299.1	<i>Python curtus</i>	-	II	<i>Python curtus</i>
<i>RD114 retrovirus</i>	ABS12072.1	RD114	<i>Gammaretrovirus</i>	I	<i>Felis catus</i>
<i>Reithrodontomys fulvescens endogenous virus</i>	HM160079.1	<i>Reithrodontomys</i>	-	-	<i>Reithrodontomys fulvescens</i>
<i>RV-Iberian frog</i>	AJ225217.1	RV-Iberian frog	-	-	<i>Iberian frog</i>
<i>RV-Slider turtleI</i>	AJ225227.1	RV-Slider turtleI	-	-	<i>Chrysemys scripta</i>
<i>RV-Tuatara</i>	AJ225236.1	RV-Tuatara	-	-	<i>Tuatara</i>
<i>Snakehead retrovirus</i>	NC001724.1	Snakehead	<i>Epsilonretrovirus</i>	I	<i>Ophicephalus striatus</i>

Apêndice A. Continuação

<b>Retrovírus</b>	<b>ID (GenBank)</b>	<b>Sigla</b>	<b>Classificação</b>	<b>Classe ERV</b>	<b>Hospedeiro</b>
<i>Transposon gypsy</i>	XP874426.1	Gypsy	-	-	<i>Bos taurus</i>
<i>unnamed protein product</i>	CAF93313.1	Tetraodon	-	-	<i>Tetraodon nigroviridis</i>
<i>Walleye dermal sarcoma virus</i>	ABO25842.1	Walleye	<i>Epsilonretrovirus</i>	I	<i>Sander vitreus</i>
<i>Walleye epidermal hyperplasia virus 2</i>	AAC59311.1	Walleye 2	<i>Epsilonretrovirus</i>	I	<i>Sander vitreus</i>
<i>Walleye epidermal hyperplasia virus type 1</i>	AF014792.1	Walleye 1	<i>Epsilonretrovirus</i>	I	<i>Sander vitreus</i>
<i>Xenopus tropicalis endogenous virus</i>	HM765512.1	XTERV1	<i>Epsilonretrovirus</i>	I	<i>Xenopus tropicalis</i>

**Apêndice B.** Lista dos clones obtidos neste estudo (classificados por espécie), seus respectivos tamanhos de fragmento, número de códons de parada e mutações e a classe de ERV a qual eles pertencem.

Espécie	Clone	Fragmento (pb)	stop codon + mutação	Classe
<i>Leopardus colocolo</i>	Lco_JO_600_8*	329	0	I
<i>Leopardus colocolo</i>	Lco_JO_600_13*	329	0	I
<i>Leopardus colocolo</i>	Lco_JO_600_14*	329	0	I
<i>Leopardus colocolo</i>	Lco_JO_1000_2	867	5	I
<i>Leopardus colocolo</i>	Lco_JO_1000_10	843	5	I
<i>Leopardus colocolo</i>	Lco_JO_1000_12	967	2	I
<i>Leopardus colocolo</i>	Lco_JO_1000_18*	295	0	I
<i>Leopardus colocolo</i>	Lco_JO_1000_20*	329	0	I
<i>Leopardus colocolo</i>	Lco_JO_1200_4*	329	0	I
<i>Leopardus colocolo</i>	Lco_JO_1200_13*	210	0	I
<i>Leopardus geoffroyi</i>	Lg_CT_1000_3	892	5	I
<i>Leopardus geoffroyi</i>	Lg_CT_1000_4	892	5	I
<i>Leopardus geoffroyi</i>	Lg_CT_1000_5	971	4	I
<i>Leopardus geoffroyi</i>	Lg_CT_1000_7	892	5	I
<i>Leopardus geoffroyi</i>	Lg_EM_1000_3	1063(584-819 SINE)‡	9†	II
<i>Leopardus geoffroyi</i>	Lg_EM_1000_5	858	5	I
<i>Leopardus geoffroyi</i>	Lg_JO_1000_1	603	3	I
<i>Leopardus geoffroyi</i>	Lg_JO_1000_2	919	8	I
<i>Leopardus geoffroyi</i>	Lg_JO_1000_6	939	0	I
<i>Leopardus geoffroyi</i>	Lg_JO_1000_7*	600	0	I
<i>Leopardus geoffroyi</i>	Lg_JO_1000_8	909	15	I
<i>Leopardus geoffroyi</i>	Lg_JO_1000_10	921	7	I
<i>Leopardus geoffroyi</i>	Lg_JO_1000_13	882	5	I
<i>Leopardus geoffroyi</i>	Lg_JO_1000_16	1245	3	I
<i>Leopardus geoffroyi</i>	Lg_JO_1000_17	895	5	I

Apêndice B. Continuação

Espécie	Clone	Fragmento (pb)	stop codon + mutação	Classe
<i>Leopardus pardalis</i>	Lp_CT_1000_5	786	3	I
<i>Leopardus pardalis</i>	LP_JO_1*	327	0	I
<i>Leopardus pardalis</i>	LP_JO_2*	326	0	I
<i>Leopardus pardalis</i>	LP_JO_3	867	4	I
<i>Leopardus pardalis</i>	LP_JO_6*	329	0	I
<i>Leopardus pardalis</i>	LP_JO_5*	326	0	I
<i>Leopardus pardalis</i>	LP_JO_4*	326	0	I
<i>Leopardus pardalis</i>	LP_JO_8*	329	0	I
<i>Leopardus pardalis</i>	LP_JO_9*	329	0	I
<i>Leopardus pardalis</i>	Lp_JO_10*	328	0	I
<i>Leopardus pardalis</i>	LP_JO_1000_2	983	6	I
<i>Leopardus pardalis</i>	LP_JO_1000_3	985	1	I
<i>Leopardus pardalis</i>	LP_JO_1000_4	982	4	I
<i>Leopardus pardalis</i>	LP_JO_1000_6	864	5	I
<i>Leopardus tigrinus</i>	Lti_JO_1000_1*	703	0	I
<i>Leopardus tigrinus</i>	Lti_JO_1000_5	882	4	I
<i>Leopardus tigrinus</i>	Lti_JO_1000_6	930	0	I
<i>Leopardus tigrinus</i>	Lti_JO_1000_9	965	6	I
<i>Leopardus tigrinus</i>	Lti_JO_1200_9	1007	1	I
<i>Leopardus wiedii</i>	Lwi_JO_700_5*	328	0	I
<i>Leopardus wiedii</i>	Lwi_JO_700_4*	589	0	I
<i>Leopardus wiedii</i>	Lwi_JO_700_7	641(314-641ERV)‡	1	I
<i>Leopardus wiedii</i>	Lwi_JO_1000_2	872	2	I
<i>Leopardus wiedii</i>	Lwi_JO_1000_3	976	2	I
<i>Leopardus wiedii</i>	Lwi_JO_1000_4	931	1	I
<i>Leopardus wiedii</i>	Lwi_JO_1200_4	930	0	I
<i>Panthera onca</i>	OC_JO_1000_1	839	4	I
<i>Panthera onca</i>	OC_JO_1000_2	985	1	I

Apêndice B. Continuação

Espécie	Clone	Fragmento (pb)	stop codon + mutação	Classe
<i>Panthera onca</i>	OC_JO_1000_5	982	2	I
<i>Panthera onca</i>	OC_CT_1000_1	933	4	I
<i>Panthera onca</i>	OC_CT_1000_2	716	1	I
<i>Puma concolor</i>	Pco_CT_700_4	601	3	I
<i>Puma concolor</i>	Pco_CT_700_7	601	3	I
<i>Puma concolor</i>	Pco_CT_1000_1	600	3	I
<i>Puma concolor</i>	Pco_CT_1000_2	282	1	I
<i>Puma concolor</i>	Pco_CT_1000_5	601	3	I
<i>Puma concolor</i>	Pco_JO_700_1	604	3	I
<i>Puma concolor</i>	Pco_JO_700_2	437	3	I
<i>Puma concolor</i>	Pco_JO_700_5	604	3	I
<i>Puma concolor</i>	Pco_JO_1000_4	929	2	I
<i>Puma concolor</i>	Pco_JO_1000_5	973	6	I
<i>Puma concolor</i>	Pco_JO_1000_6	1210(168-399 SINE)‡	1†	I
<i>Puma concolor</i>	Pco_JO_1000_7	971	6	I
<i>Puma concolor</i>	Pco_JO_1000_9	982	4	I
<i>Puma concolor</i>	Pco_JO_1200_1	1018	3	I
<i>Puma concolor</i>	Pco_JO_1200_3	1018	3	I
<i>Puma concolor</i>	Pco_JO_1200_4	953	4	I
<i>Puma yagouaroundi</i>	PY_CT_1000_1	979	2	I
<i>Puma yagouaroundi</i>	PY_CT_1000_2	985	5	I
<i>Puma yagouaroundi</i>	PY_CT_700_5	594	1	I
<i>Puma yagouaroundi</i>	PY_EM_1300_1	695	13	II
<i>Puma yagouaroundi</i>	PY_EM_600_2	500	11	II
<i>Puma yagouaroundi</i>	Py_EM_1300_3	695	13	II
<i>Puma yagouaroundi</i>	PY_JO_1000_1	971	2	I
<i>Puma yagouaroundi</i>	PY_JO_500_1	602	2	I

Apêndice B. Continuação

Espécie	Clone	Fragmento (pb)	stop codon + mutação	Classe
<i>Puma yagouaroundi</i>	PY_JO_600_4	599	1	I
<i>Puma yagouaroundi</i>	PY_JO_1000_2	971	2	I
<i>Puma yagouaroundi</i>	PY_JO_1000_3	928	3	I
<i>Puma yagouaroundi</i>	PY_JO_1000_4	984	5	I
<i>Puma yagouaroundi</i>	PY_JO_1000_5	602	1	I

(\*) ORF parcial sem presença de *stop codons* e mutações que ocasionam a mudança da fase de leitura, mas com deleção da PR e parte da RT, portanto, foram consideradas defectivas; (†) Não computado na sequência do SINE; (‡) Posição na sequência.



**Apêndice C.** Lista dos SINEs utilizados na filogenia da Figura 4. Estas sequências foram obtidas no NCBI através do programa BLASTn utilizando o banco de dados nr. A sequência SINEC Fc2 foi obtida utilizando o programa BLAT.

<b>Sine</b>	<b>Espécie</b>	<b>Gene/Localização</b>	<b>Identificação</b>
<i>F. bieti_Zfy</i>	<i>Felis. bieti</i>	<i>Zfy gene (zink finger protein -chrY)</i>	AY518634 .1
<i>F. catus_Zfy</i>	<i>Felis. catus</i>	<i>Zfy gene (zink finger protein -chrY)</i>	AY518632.1
<i>F. catus_AGPAT1_MHC</i>	<i>Felis. catus</i>	<i>MHC-1-acylglycerol-3-phosphate O-acyltransferase 1</i>	EU153401.1
<i>F. catus_APOBECC3H</i>	<i>Felis. catus</i>	<i>Felis catus antiviral cytidine deaminase (APOBECC3H)</i>	EU109281
<i>F. catus_ASIP</i>	<i>Felis. catus</i>	<i>Agouti signaling protein (ASIP)</i>	AY237394.1
<i>F. catus_AY152827_MHC</i>	<i>Felis. catus</i>	<i>MHC-Class II</i>	AY152827.1
<i>F. catus_DPBp_MHC</i>	<i>Felis. catus</i>	<i>MHC-Class II antigen beta chain, pseudogene</i>	EU153401.1
<i>F. catus_MHC_X11</i>	<i>Felis catus</i>	<i>FLA distal class I -X11</i>	EU153402.1
<i>F. catus_SMN_sine1</i>	<i>Felis catus</i>	<i>survival of motor neuron (SMN) gene</i>	AF503618.1
<i>F. catus_SMN_sine2</i>	<i>Felis catus</i>	<i>survival of motor neuron (SMN) gene</i>	AF503618.1
<i>F. libica_zfy</i>	<i>Felis.libica</i>	<i>Zfy gene (zink finger protein -chrY)</i>	AY518635.1
<i>F. margarita_Zfy</i>	<i>Felis margarita</i>	<i>Zfy gene (zink finger protein -chrY)</i>	AY518636.1
<i>F. silvestris_Zfy</i>	<i>Felis silvestris</i>	<i>Zfy gene (zink finger protein -chrY)</i>	AY518633.1
<i>N. nebulosa_ AC159158</i>	<i>Neofelis nebulosa</i>	Desconhecido	AC159158.4
<i>N. nebulosa_ AC152898</i>	<i>Neofelis nebulosa</i>	Desconhecido	AC152898.3

Apêndice C. Continuação

Sine	Espécie	Gene/Localização	Identificação
<i>F. bieti_Zfy</i>	<i>Felis. bieti</i>	<i>Zfy gene (zink finger protein -chrY)</i>	AY518634 .1
<i>F. catus_Zfy</i>	<i>Felis. catus</i>	<i>Zfy gene (zink finger protein -chrY)</i>	AY518632.1
<i>N. nebulosa_AC158253</i>	<i>Neofelis nebulosa</i>	Desconhecido	AC158253.3
<i>N. nebulosa_AC162123</i>	<i>Neofelis nebulosa</i>	Desconhecido	AC162123.3
<i>O. manul_Ube1y</i>	<i>Otocolobus manul</i>	<i>Ube1y gene (ubiquitin activating enzyme E1- chrY)</i>	AY518631.1
<i>P. tigris</i>	<i>Panthera tigris altaica</i>	Desconhecido	GU071272.1
SINEC_Fc2	<i>Felis. catus</i>	chrX:61882383-61882587 *	‡

(\*) Posição no genoma do gato doméstico; (‡) montagem versão catChrV17e do genoma do gato doméstico produzida por NHGRI Genome Technology Branch and NIH Intramural Sequencing Center e disponível em <http://genome.ucsc.edu>

**Apêndice D.** Alinhamento de PyEM13003 e representantes de retroelementos, conforme descrito na Figura 15 . Box 1-4 representam os respectivos domínios da transcriptase reversa.

	1	2	3
	*	20	*
	*	40	*
	*	60	*
	*	80	*
	*	100	*
	*	120	*
PyEM13003	: PCLFEQWPRKGEKSKQTKNLGLXLLLEASHIESS--TSPWNIPIFTIPKKS-G-K-WRLLHDLRVINAIQSMGALKARXCPTLLFLDQEX-PI---DLKDRFXTIPIP----L-EKDKECFASFLLV	: 111	
HERVK22	: FVWVXGSIKKEKLEHIQCLVQEQQLDAGHIEPT--TSPWNTPIFTIPKRS-G-K-WRLLYDLHAINAVMFPXGTFXXRXCKSPVMIPKDW-PLIVIDLKDCFLXXXXXXXXLHPEDXDKFAFTIPT	: 119	
Casuaris	: FVWVNWQPLKAERLQKLHGLVQEQQLCACHIVPS--TSPWNTPVFVIKPKNG-K-WRLLHDLRAVNALIEPMGTLQPGXXXXXXXXLCKPSSW-PLVVIDLKDCFFTIPIP----LHPADAPRFVPA	: 115	
Reithrodon	: FVWVQWPLTTEKQLALEELVEQLNAQHIESS--TSPWNSPVFVIKKS-G-K-WRNVTDLRAINKVIQPMGSLQSNXKCLCCXKGW-PLIVIDLKDCFFSIPIP----LQEKDRERFAFTVPT	: 115	
Nyctomys	: FVWVQWPLTSEKQLALEELVQEQQLNAQHIESS--TSPWNSPVFVVKKS-G-K-WRNVTDLRAVNKVIQPPGSLQSNXKCLCCXKGW-PLIVIDLKDCFFTIPIP----LQEKDRERFAFTVPT	: 115	
Oryzomys	: FIWIIPQWPLTKEKQLALEELVQEQQLNAQHIESS--TSPWNSPVFVIKKS-G-K-WRNVTDLRAINKVIQPMGSLQSNXKCLCCXKGW-PIIVIDLKDCFFTIPIP----LQEKDRERFAFTVPT	: 115	
MMTV	: FVWVQWPLTQEKQLALEELVQEQQLNAQHIESS--NSPWNTPVFVIKKS-G-K-WRLLQDLRAVNATMHDGALQPG-LPTPSAIPDKS-YIIIIIDLKDCFFNIK----LHPEDCKRFVPAFTVPT	: 114	
Jaagsiekte	: FVWVQWPLTQEKLSAPQQLVQEQQLRGLHIEPS--TSAWNSPIFVIKKS-G-K-WRLLQDLRAVNATMHDGALQPG-LPTPSAIPDKS-YIIIIIDLKDCFFTIPIP----LAPQDYKRFVPAFTVPT	: 114	
ALV	: FVWIDQWPLTQEKLSAPQQLVQEQQLRGLHIEPS--LSCWNTPVFVIKKS-G-S-YRLLHDLRAVNATMHDGALQPG-LPTPSAIPDKS-YIIIIIDLKDCFFSIPIP----LAEQDYKRFVPAFTVPT	: 114	
BLV	: XXXXXXPFKLERLQALQDLVHRSLEAGYISFW--DGPGNPVFVVRKENG-A-WRFVHDLRVNALTKEPIALSPG-PDDLTAIPDKS-YIIIIIDLKDCFFTIPIP----VEDRFRSYFAFTLPT	: 115	
FeLV	: FISIROYPMPEAYQGIKPHIRMLDQGLKPC--QSPWNTPLLVKPKGTDK-YRQVQDLREVNKRVEDIHPVNPYNLLSTLPSHPWYTVLDDKDAFFCLR----LHSESQLLFVPAFTVPT	: 117	
Walleye	: XPSIRQYPLPKDTEGLRPLISLENGELIKC--HSPCNTPIFPIKAGRDE-YRMIHDLRAINNI VAPLTAVASPTTVLSNLAPSLHWFTVIDLSNAFFSVP----LHSDSXYLFAFTVPT	: 117	
Gypsy	: XXXXXXPIPRGARI GILPHISRLKQAGLLVBC--QTAWNTPIPVKKEG-QXYRFPVQDLRLVQATVTLHLSVVKPYTLLSLLPKTRIYTCCLDLTEAFSRIC----LAPASQPIFAFTVPT	: 117	
FSV	: ETPQKQYHINPKAKEDIQIVINDLLKQGLIQK--ESTMNTPVYVVKENG-R-WRNVLDRAVNKVTPLIAVQNHYSYGLIGSLFKGK-YKTVIDLSNGFWAHP----LVPEDYWITFAFTVPT	: 115	
FIV	: GPQIKQWPLTNEKIEALTEIVREREREKVKRADENNPWNTPVFAIKKS-G-K-WRMLIDFRELNKQTEDLTAQGLG-LPHPGGLQKK-QVTVLDIGDAYETIPIP----LDPDYAPYTAFTLPT	: 116	
CAEV	: GPHVQWPLTBEKLGKLTETIDKLVBEGLKQAPPHWTCNTPIFCKIKKS-G-K-WRMLIDFRELNKQTEDLTAQGLG-LPHPGGLQKK-QVTVLDIGDAYETIPIP----LYEPYREYTCFTLLS	: 116	
	*	140	*
	*	160	*
PyEM13003	: TNSQEPISRYQWMTMSRXXOGMINSPTICRYFIQOALLEPLKLGWIKSY	: 158	
HERVK22	: YNNQQPVERYQWTVLP--QGMMNSPTICQLYVHEALLPVCQSFQAKIF	: 166	
Casuaris	: VNNQESTLRFHWTVLP--QGMLNSPTICQMFVARAILPIRRAYPNALIY	: 162	
Reithrodon	: YNNSQPAKRYQWRVLP--QGMLNSPTLCQYFVQQPLEEIRKKEFKSIY	: 162	
Nyctomys	: YNNSQPVRRFHWTVLP--QGMLNSPTLCQYFVQQPLEEIRKKEFKSIY	: 162	
Oryzomys	: YNNSQPQKRYQWRVLP--QGMLNSPTLCQYFVQKPLEEIRKKEFKSIY	: 162	
MMTV	: ENFKRPQRQFQWVLP--QGMLNSPTLCQKQFVDAKIALTVRKYQDFYIV	: 161	
Jaagsiekte	: VNFKEPMORYQWRVLP--QGMTNSPMLCQKQFVATALAPVQRFPQLYLV	: 161	
ALV	: VNNQAPARFQWVLP--QGMTCSPTICQLIVGQVLEPLRLKHPSLCML	: 161	
BLV	: FGGQLPHRRFAWRVLP--QGFINS PALFERALQELPRQVSAASFQSLLV	: 162	
FeLV	: EXXXXX-QLTWTRLP--QGFKNSTPLFDEALHSDLADFRVRYPALVLL	: 163	
Walleye	: -----HQYTWTVLP--QGFHNSPTLFSQALYQSLHKIKFKISSIICI	: 157	
Gypsy	: EX-GGNKQLTWTHTLS--QGFKNSTPLFDEALHSDLEPFQPERYGCCLL	: 163	
FSV	: -----KQYCWTVLP--QGFINS PGLFTGDVVDLLOXX----PNVVEY	: 151	
FIV	: KNNAGPGRFVWCSLP--QGWILSPLIYQSTLDNIIQPFIRQNPQLDIY	: 163	
CAEV	: ENNLGPCKRYWVLP--QGWKLSPSVYQFTMQEILEDWIOQHPKIXXX	: 163	

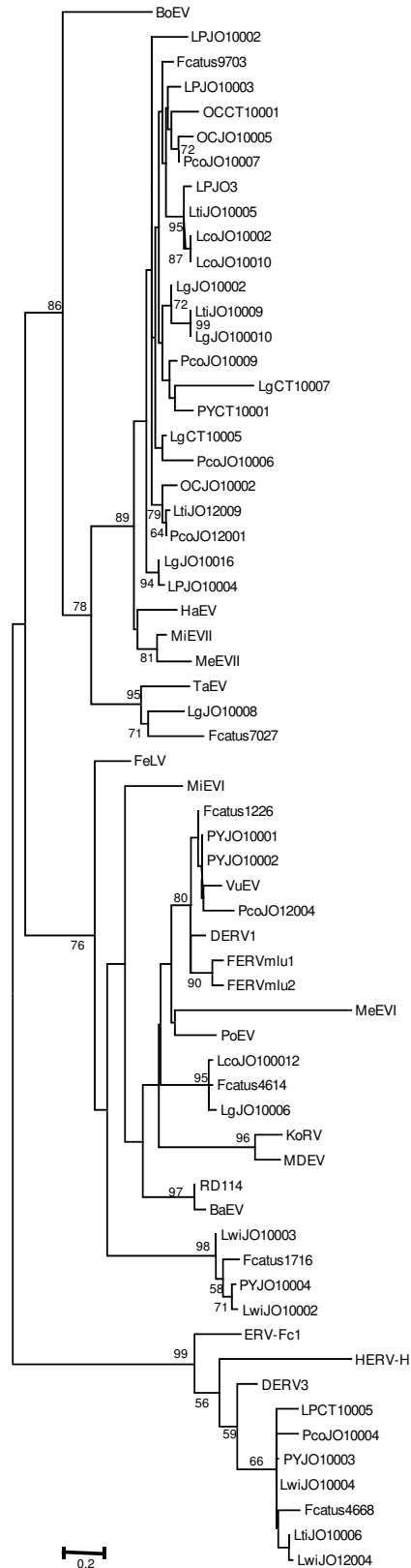
**Apêndice E.** Lista das sequências obtidas através de buscas (com programa tBLASTn no banco de dados WGS) de retroelementos similares as sequências (*query*) do grupo G1 (PyJO10001, PyJO10004, LgJO10006), grupo G2 (LcoJO10002, OCCT10001, LgJO100010 LgJO10008) e grupo G3 (LtiJO10006). A lista representa apenas a sequências do banco (*hit*) que obteve maior escore dentro de cada espécie, em cada busca.

Grupo G1		Grupo G2		Grupo G3	
Especie	N°(GenBank)	Especie	N°(GenBank)	Especie	N°(GenBank)
<i>Callithrix jacchus</i>	ACFV01195862	<i>Ailuropoda melanoleuca</i>	ACTA01103996	<i>Callithrix jacchus</i>	ACFV01199700
<i>Canis lupus familiaris</i>	AAEX03011462	<i>Bos taurus</i>	DAAA02054755	<i>Canis lupus familiaris</i>	AAEX03000317
<i>Choloepus hoffmanni</i>	ABVD01746698	<i>Callithrix jacchus</i>	ACFV01076043	<i>Dipodomys ordii</i>	ABRO01327136
<i>Echinops telfairi</i>	AAIY01041058	<i>Dasyopus novemcinctus</i>	AAGV020988216	<i>Felis catus</i>	ACBE01494668
<i>Erinaceus europaeus</i>	AANN01599932	<i>Felis catus</i>	AANG02197027	<i>Gorilla gorilla</i>	CABD02403795
<i>Felis catus</i>	AANG02041226	<i>Felis catus</i>	ACBE01279703	<i>Homo sapiens</i>	AADB02013597
<i>Felis catus</i>	ACBE01041716	<i>Gorilla gorilla</i>	CABD02148830	<i>Macaca fascicularis</i>	CAEC01255977
<i>Felis catus</i>	AANG02174614	<i>Heterocephalus glaber</i>	AFSB01203456	<i>Macaca mulatta</i>	AANU01007707
<i>Gorilla gorilla</i>	CABD02454890	<i>Homo sapiens</i>	AADB02036889	<i>Microcebus murinus</i>	ABDC01155485
<i>Heterocephalus glaber</i>	AFSB01151231	<i>Macaca mulatta</i>	AANU01113705	<i>Mustela putorius furo</i>	AEYP01092408
<i>Macaca fascicularis</i>	AEHL01391779	<i>Macaca mulatta</i>	AANU01214376	<i>Saimiri boliviensis</i>	AGCE01123977
<i>Macaca fascicularis</i>	CAEC01304422	<i>Microcebus murinus</i>	ABDC01264851	<i>Tarsius syrichta</i>	ABRT010187673
<i>Macaca mulatta</i>	NW001196508	<i>Myotis lucifugus</i>	AAPE02035021	<i>Tursiops truncatus</i>	ABRN01231801

Apêndice E. Continuação

Grupo G1		Grupo G2		Grupo G3
<i>Macaca mulatta</i>	AANU01062806	<i>Myotis lucifugus</i>	AAPE02037011	
<i>Mus musculus</i>	AAHY01561886	<i>Myotis lucifugus</i>	AAPE02067297	
<i>Myotis lucifugus</i>	AAPE02049384	<i>Nomascus leucogenys</i>	ADFV01123118	
<i>Myotis lucifugus</i>	AAPE02002964	<i>Otolemur garnettii</i>	AAQR03105487	
<i>Ochotona princeps</i>	AAYZ01197507	<i>Pan troglodytes</i>	AADA01173210	
<i>Oryctolagus cuniculus</i>	AAGW02065556	<i>Pongo abelii</i>	ABGA01038815	
<i>Pan troglodytes</i>	AACZ03156450	<i>Pongo abelii</i>	ABGA01076357	
<i>Procavia capensis</i>	ABRQ01832156	<i>Procavia capensis</i>	ABRQ01348879	
<i>Pteropus vampyrus</i>	ABRP01299030	<i>Pteropus vampyrus</i>	ABRP01035452	
<i>Rattus norvegicus</i>	AAHX01271931	<i>Saimiri boliviensis</i>	AGCE01132573	
<i>Spermophilus tridecemlineatus</i>	AAQQ01194141	<i>Tarsius syrichta</i>	ABRT010331818	
<i>Tarsius syrichta</i>	ABRT010822206	<i>Tarsius syrichta</i>	ABRT010400171	
<i>Tursiops truncatus</i>	ABRN01395693	<i>Vicugna pacos</i>	ABRR01535793	
<i>Tursiops truncatus</i>	ABRN01209221	<i>Vicugna pacos</i>	ABRR01535793	

**Apêndice F.** Árvore de Máxima Verossimilhança baseada em um alinhamento de 68 aminoácidos da região da protease e modelo JTT+G. Os números próximos aos ramos representam os valores de bootstrap > 50%. Lista dos clones encontra-se no Apêndice B, demais ERVs no Apêndice A e sequências de *Felis catus* no Apêndice E.



**Apêndice G.** Árvore de Máxima Verossimilhança baseada em um alinhamento de 170 aminoácidos da região da transcriptase reversa e modelo JTT+G. Os números próximos aos ramos representam os valores de bootstrap > 50%. Lista dos clones encontra-se no Apêndice B, demais ERVs no Apêndice A e as sequências de *Felis catus* no Apêndice E.

