RETROVIROLOGY

SHORT REPORT Open Access

Endogenous lentivirus in Malayan colugo (*Galeopterus variegatus*), a close relative of primates

Tomáš Hron[†], Helena Fábryová[†], Jan Pačes[*] and Daniel Elleder[*]

Abstract

Background: A significant fraction of mammalian genomes is composed of endogenous retroviral (ERV) sequences that are formed by germline infiltration of various retroviruses. In contrast to other retroviral genera, lentiviruses only rarely form ERV copies. We performed a computational search aimed at identification of novel endogenous lentiviruses in vertebrate genomes.

Findings: Using the *in silico* strategy, we have screened 104 publicly available vertebrate genomes for the presence of endogenous lentivirus sequences. In addition to the previously described cases, the search revealed the presence of endogenous lentivirus in the genome of Malayan colugo (*Galeopterus variegatus*). At least three complete copies of this virus, denoted ELVgv, were detected in the colugo genome, and approximately one hundred solo LTR sequences. The assembled consensus sequence of ELVgv had typical lentivirus genome organization including three predicted accessory genes. Phylogenetic analysis placed this virus as a distinct subgroup within the lentivirus genus. The time of insertion into the dermopteran lineage was estimated to be more than thirteen million years ago.

Conclusions: We report the discovery of the first endogenous lentivirus in the mammalian order Dermoptera, which is a taxon close to the Primates. Lentiviruses have infiltrated the mammalian germline several times across millions of years. The colugo virus described here represents possibly the oldest documented endogenization event and its discovery can lead to new insights into lentivirus evolution. This is also the first report of an endogenous lentivirus in an Asian mammal, indicating a long-term presence of this retrovirus family in Asian continent.

Keywords: Endogenous lentiviruses, Dermoptera, Paleovirology

Findings

The lentiviruses have been described in several mammalian orders, including Primates, Artiodactyls, Perissodactyls, and Carnivores. They are the cause of a variety of chronic diseases and constitute a major public health concern, especially due to the HIV/AIDS pandemic. In contrast to other retroviral genera, lentiviruses rarely generate ERV copies [1]. The ERVs are formed following germline infection and further vertical transmission of the integrated provirus [2]. The presence of such genomic "viral fossils" enables the study of long-term evolutionary history and evolution of lentiviruses [1]. The first endogenous lentivirus has been described in 2007 in the genome of European rabbit [3]. Since then, there have been only a few additional reports of lentiviruses infiltrating into the genomes of hares, lemurs and ferrets [4-8]. We have performed a large-scale screening of all publicly available vertebrate genomes for the presence of endogenous lentivirus sequences. Here, we report the identification of the first endogenous lentivirus in the mammalian order Dermoptera, in the genome of the Malayan colugo (*G. variegatus*). We discuss the genomic and phylogenetic characteristics of this virus, which place it as one of the oldest described members of the lentivirus genus.

We have implemented a computational approach based on automated BLAST searches and the best bidirectional hit (BBH) strategy against custom retroviral database. This enabled us to screen for candidate lentiviral sequences in multiple genomic datasets (Figure 1A).

* Correspondence: jan.paces@img.cas.cz; daniel.elleder@img.cas.cz
[†]Equal contributors
Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Vídeňská 1083, 14220 Prague, Czech Republic

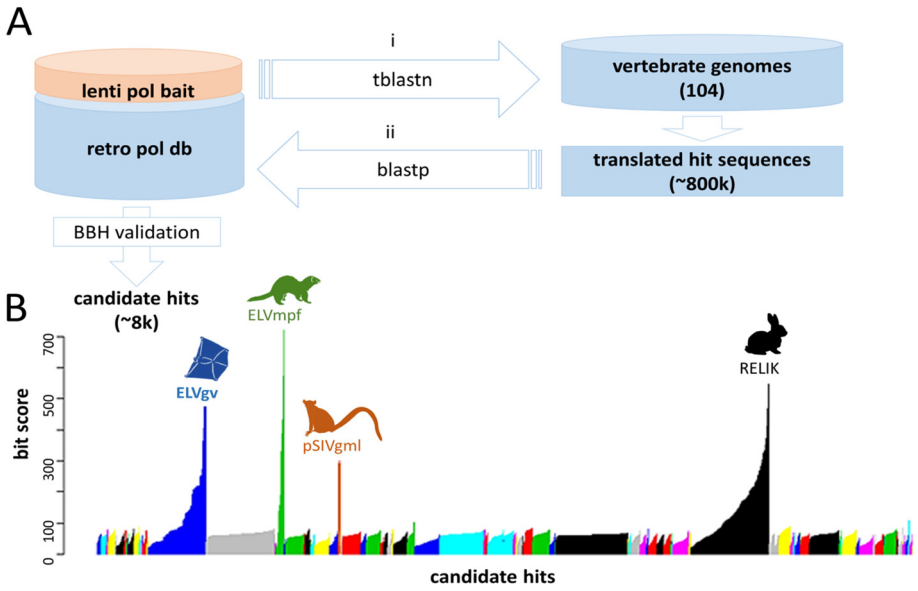

Figure 1 Screening for lentiviral ERVs. (A) Schematic depiction of the computational screening pipeline. The first step of the best bidirectional hit (BBH) strategy was performed by tblastn search in vertebrate genome database to identify candidate endogenous lentivirus fragments (i). In this step the following Pol amino acid sequences were used as baits: human immunodeficiency virus type 1 (HIV-1), feline immunodeficiency virus (FIV), Visna/maedi virus, rabbit endogenous lentivirus type K (RELIK), gray mouse lemur prosimian immunodeficiency virus (pSIVgml), and domestic ferret (*Mustela putorius furo*) endogenous lentivirus (ELVmpf). The cutoff for the blast search was set at E-value < 10^{-5}. To filter out non-lentiviral sequences, translated hits were used as a query for backward blastp search against database of retroviral Pol sequences belonging to all retroviral genera (ii). Hits aligned with the best bit score to lentiviral sequences in the backward blast search were further analyzed. **(B)**. Graph shows bit scores of all lentiviral candidate hits ordered by species in which they were found. Each species is represented by different color. Newly discovered lentiviral sequence in colugo (ELVgv) as well endogenous lentiviruses in rabbit [3], domestic ferret [6], and gray mouse lemur [4,5] are indicated. Previously published endogenous lentivirus sequences were excluded as baits for their corresponding host species (e.g. RELIK against the rabbit genome) to avoid identical matching of the hits.

A search of 104 publicly available vertebrate genomes recovered 8,179 candidate hits, each aligned to lentiviral sequence with a given bit score (Figure 1B). We identified false positive bit scores <100 in majority of animals. However, a few hits from rabbit, domestic ferret, and grey mouse lemur reached significantly higher bit scores. These sequences corresponded to previously described endogenous lentiviruses in the above mentioned species [3-6,8]. High scoring hits were also found in the genome of colugo. The matching sequences were manually extracted and found to cluster robustly with lentiviruses upon preliminary phylogenetic analysis. The endogenous lentivirus in the *G. variegatus* genome was denoted ELVgv.

Further BLAST searches of the colugo genomic contigs revealed the presence of three complete ELVgv proviruses (provirus *I* at positions 11,594-19,841 of contig JMZW01084956; provirus *II* at positions 14,164-23,469 of contig JMZW01174031; provirus *III* at positions 40,701-51,516 of contig JMZW01021293). This search also identified approximately 100 solo long terminal repeats (LTR), which are formed by recombination between the two LTRs flanking the viral internal sequences [9]. The BLASTn parameters employed for the identification of solo LTRs were the following: e-value < 10^{-100}, identity to the LTR of full-length ELVgv provirus at least 80%, and coverage at least 50%. In addition, several smaller contigs containing fragments of internal virus sequences were detected (data not shown). The colugo genome assembly covers majority of the genome (assembly size 2.8 Gbp, accession number JMZW00000000), therefore it can be assumed that there are at least three complete provirus copies and ~30 times more solo LTRs per genome.

Alignment of all available contig sequences was used to reconstruct the ELVgv full consensus sequence (Figure 2 and Additional files 1 and 2). The reconstructed provirus is 10,040 bp long and flanked by LTRs of approximately 420 bp. The genome organization is typical for a lentivirus, with three long open reading frames (ORFs) corresponding to *gag*, *pol*, and *env* genes. The *gag* and *pol* genes lie in different reading frames and *pol* is predicted to be translated via ribosomal frameshifting. Consistent with that, a hairpin RNA secondary structure is predicted in the *gag-pol* overlapping region (Additional file 3) [10]. A feature present also in other nonprimate lentiviruses is the occurrence of dUTPase between RNaseH and integrase domains of the ELVgv *pol* gene. Two short ORFs, named

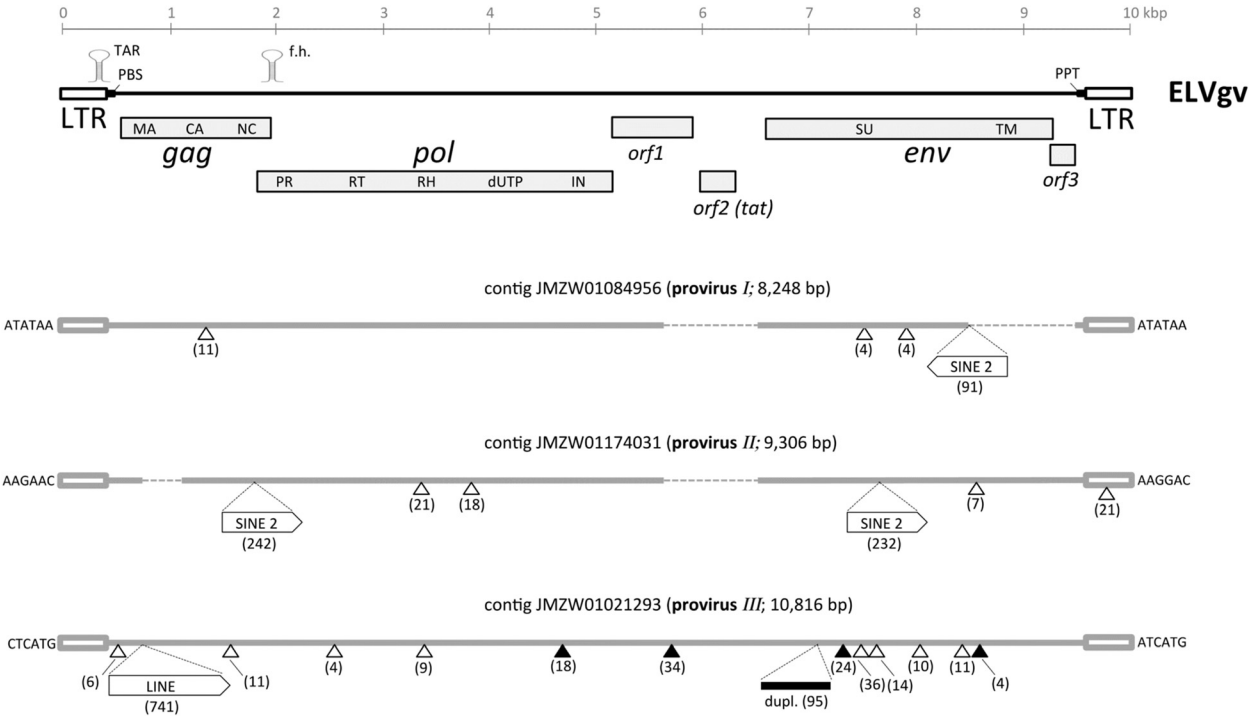

Figure 2 Genome organization of ELVgv. The consensus sequence of ELVgv is shown schematically below the scale. The position of open reading frames and other genomic features is indicated. The structure of the three complete proviruses recovered from the whole genome shotgun (WGS) contigs is shown below; their accession numbers are [GenBank:JMZW01084956], [Genbank:JMZW01174031], and [Genbank: JMZW01021293]. For each contig, the length of the provirus and its corresponding target site duplication is shown. Deletions and insertions >3 bp are depicted by open and closed triangles, respectively, together with their length. The four longest deletions are indicated by thin dashed line. Insertions of retroelements were detected by Censor [11] and are depicted by large open arrows, drawn not to scale. ENV coding regions were characterized with the help of specialized prediction servers [12-14]. LTR, long terminal repeat; MA, matrix; CA, capsid; NC, nucleocapsid; PR, protease; RT, reverse transcriptase; RH, RNaseH; dUTP, dUTPase; IN, integrase; PBS, primer binding site; f.h., frameshift hairpin; PPT, polypurine tract; SINE, short interspersed nuclear element; LINE, long interspersed nuclear element; dupl., duplication.

orf1 and *orf2*, were detected in the intervening region between *pol* and *env* (Figure 2 and Additional file 1). The *orf2* (103 aa) could be identified by sequence similarity as a *tat* gene (Additional file 3). A corresponding TAR (transactivating responsive region) was predicted in the LTR downstream of the putative promoter (Additional file 3) [10]. The *orf1* gene (272 aa) partially overlaps in an alternative reading frame with *pol*. No sequence similarity of *orf1* with any lentiviral accessory gene was detected. According to its size and genomic location, *orf1* might encode a *vif* homolog. A third short ORF, *orf3* (83aa), partially overlaps with the end of *env*, and extends towards the 3'LTR. As for *orf1*, the sequence of *orf3* did not point to any specific accessory gene. The location and size indicate that *orf3* might be a homolog of lentiviral *rev*. The presence of a limited number of viral accessory genes is in agreement with the previously described evolutionarily ancient lentiviruses [4].

To establish the phylogenetic placement of ELVgv within lentiviruses, we have aligned the amino acid sequence of the highly conserved reverse transcriptase (RT) region of *pol* with sequences from representatives of all retrovirus genera. In subsequent phylogenetic analysis using both maximum likelihood (ML) and Bayesian methods, ELVgv RT clustered inside the lentivirus clade with high support (ML bootstrap 100, Bayesian posterior probability = 1) (Figure 3A; alignment is available in Additional file 4). In accordance with this clustering, the highest-scoring BLASTp hits of ELVgv *gag*, *pol* and *env* genes were the genes from a lentivirus, feline immunodeficiency virus (FIV; the similarity/identity to FIV counterparts of *gag*, *pol* and *env* genes were 48%/31%, 54%/35% and 27%/17%, respectively). To analyze the relationship of ELVgv to other lentiviruses in more detail, we have used the dataset of conserved regions of *gag* and *pol* lentiviral sequences from Gilbert *et al*. [5], together with the recently described ELVmpf [6,8]. ML phylogenies generated using this alignment placed ELVgv as a deep branch of the lentivirus tree (Figure 3B; alignment is available in Additional file 5), forming a distinct lentivirus subgroup. As in previous analyses of lentivirus phylogenies, basal nodes did not have strong support [3-6], and the ML tree differed slightly from the phylogeny obtained by Bayesian analysis (compare Figure 3B

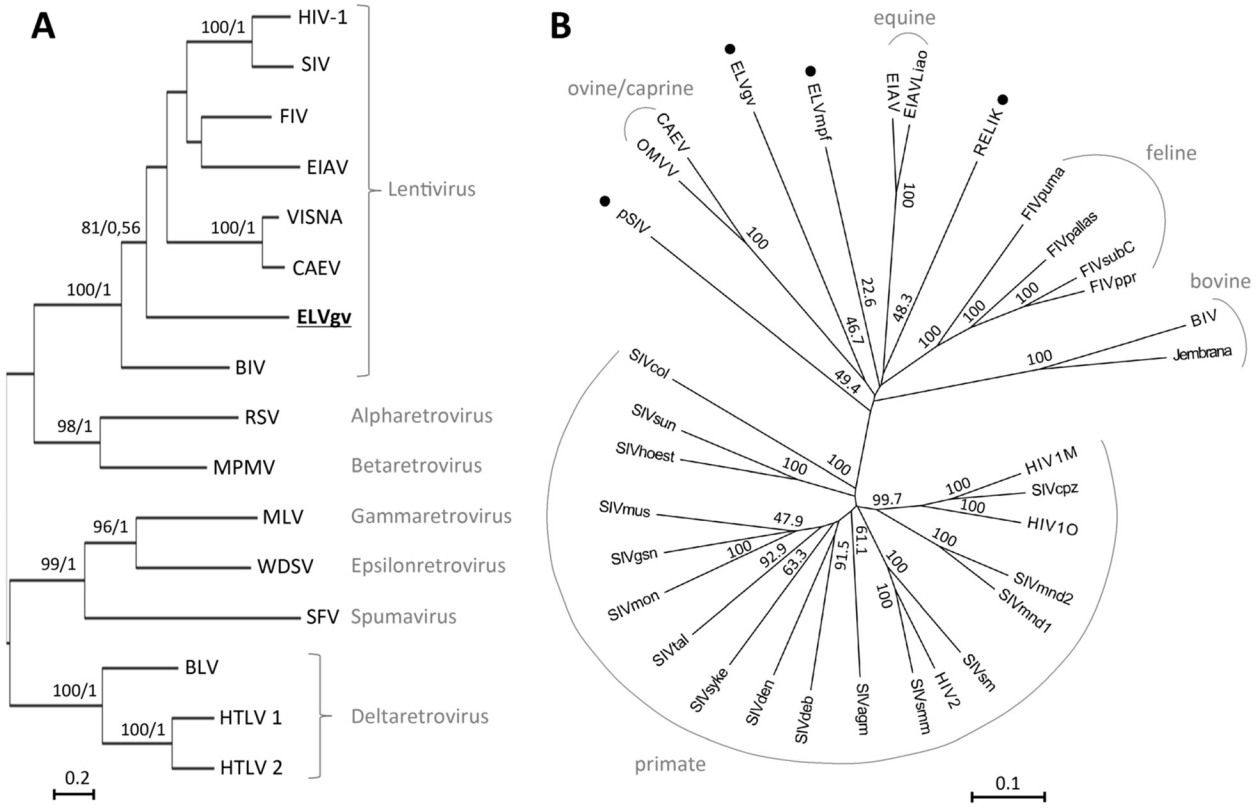

Figure 3 Phylogenetic relationship of ELVgv to other retroviruses. (A) Phylogeny of ELVgv and other retroviruses, based on alignment of RT amino acid sequences (Additional file 4 contains the alignment in FASTA format and the full names of the retroviruses). The alignment was generated in MEGA5 software [15] using the MUSCLE algorithm [16]. The ML tree was constructed in MEGA5 software, using the rtREV amino acid substitution matrix [17], Nearest-Neighbor-Interchange ML heuristic method and otherwise default parameters. Support for ML tree was assessed by 1,000 nonparametric bootstrap replicates. Bayesian analysis was run for 200,000 steps, sampling every 1,000 steps and discarding first 25% of the trees. Average standard deviation of split frequencies converged during 10,000 steps bellow 0.001. The amino acid model F81 in program MrBayes was used [18]. The support values are indicated above the branches (percent Bootstrap scores/Bayesian posterior probabilities). The ELVgv position is highlighted in bold and underlined. The names of the retrovirus genera are shown on the right. **(B)** Phylogenetic relationship of ELVgv to other exogenous and endogenous lentiviruses. The analysis was based on alignment including 2,350 most conserved nucleotides of *gag-pol* from 31 lentiviruses (used also in Gilbert *et al.* [5]), together with ELVmpf [6,8], and ELVgv from this work (alignment available in Additional file 5). The alignment was generated in MEGA5 program [15] using the MUSCLE algorithm [16]. The ML analysis was performed using MEGA5 program under Tamura-Nei model, Nearest-Neighbor-Interchange ML heuristic method and otherwise default parameters. Bootstrap supports (percent out of 1,000 replicates) are shown. Grey lines designate groups of exogenous lentiviruses. The endogenous lentiviruses are denoted by black dots. Scale bars indicate number of substitutions (amino acid or nucleotide) per site.

and Additional file 6). While in the ML analysis ELVgv clustered weakly (bootstrap support 46.7) together with the ovine/caprine lentivirus subgroup, in the Bayesian tree it formed an isolated deep branch. Separate analysis of the *gag* and *pol* genes excluded any evident recombination event (data not shown). Re-running the analysis with the three individual provirus sequences instead of the reconstructed ELVgv consensus sequence also did not influence the results (ML tree in Additional file 7). Therefore, the precise relationship of ELVgv to primate and nonprimate lentivirus groups could not be determined.

There are four lines of evidence suggesting that ELVgv inserted into the colugo germline millions of years ago. First, the three complete proviruses accumulated many genetic defects. These include insertions and deletions of various sizes, multiple frameshifts and stop codons, and insertions of SINE and LINE sequences (Figure 2). Second, the solo LTRs are formed only after prolonged existence in the germline [9]. Third, comparison of LTR sequences belonging to individual proviruses can be used to estimate the insertion times [19]. These estimates are only very approximate and use the fact that the 5' and 3' LTRs are identical at the time of insertion. Any divergence between them is supposed to have been formed postintegration and at neutral substitution rate of the host genome [19]. We assumed the range of mammalian substitution rates to be between 2.2 and 4.5 × 10^{-9} per site per year [20,21]. The provirus *I* had 20 differences between 5' and 3' LTRs, resulting in an estimated time of insertion of 5.1 - 10.3 million years ago (MYA). Similarly,

proviruses *II* and *III* yielded integration time estimates of 10.1 - 20.7 MYA and 13.2 - 27.0 MYA, respectively. We note that all three proviruses have different perfect or almost perfect target site duplications, indicating that they have not undergone recombination events after integration and that the LTRs belong to the original integrating virus (Figure 2). The genetic distances between the individual proviruses are between 0.078 and 0.105 substitutions per site. However, we did not attempt to use the distances to estimate the integration age. It is not known whether they were formed by independent insertions of circulating exogenous virus, by reinfection of germline cells or by intracellular retrotranspositions. In addition, the assembly of genomic contigs from short Illumina reads is inherently very difficult in repeat regions that include ERVs. Especially the parsing of reads among the orthologous internal positions of different proviruses might not be exact. A fourth line of evidence pointing to ancient origin of ELVgv came from the fact that seven of the solo LTR insertions reside in regions of apparent segmental genomic duplications (Additional file 8). The virus integration must have happened before the duplication event. This allows estimating the lower age limit of the integrations, which is up to 7 MYA.

The Malayan colugo (*G. variegatus*) belongs to a tiny order Dermoptera, which contains only one additional extant species, Philippine colugo (*Cynocephalus volans*) [22]. Colugos, primates, and treeshrews (Scandentia) cluster together in a taxonomic subgroup Euarchonta [23]. There is an ongoing dispute about the placement of Dermoptera. Chromosome painting comparison of these groups suggested that tree-shrews and colugos had a closer phylogenetic relationship and formed a sister group to primates [24]. However, screening of protein-coding exons indicated that colugos are closer to primates than to tree-shrews [25]. In either scenario, the split of the dermopteran lineage is estimated to be between 80–90 MYA. This is considerably older than the highest estimate of the ELVgv insertion age and indicates that the genome invasion was an independent event in Dermoptera. In accordance with this fact, about half of the ELVgv integration sites could be identified in primates and other mammals in its empty pre-integration form (data not shown). It will be informative to ascertain the presence of ELVgv in the *Cynocephalus* genus, which diverged from the genus *Galeopterus* about 18.3 MYA [25,26], and in the multiple subspecies of *Galeopterus variegatus* [22]. The timescale of the ELVgv genome infiltration is at the upper limit of the previously described lentiviral invasions in leporid species (12 MYA) [3,7], lemurs (4.2 MYA) [4,5] and ferrets (12 MYA) [6,8]. The source and ancestral relationships between these ancient lentiviruses are not possible to resolve with the current data due to the inconclusive nature of phylogenetic analyses. The ancient origin and presence in a potentially closest relative of primates makes the colugo virus an interesting addition to the lentivirus family and may add to our understanding of lentivirus evolution.

Additional files

Additional file 1: ELVgv consensus sequence with annotation. The positions of individual virus genes and their domains were determined by alignments with other lentiviral genomes. In the *env* gene, the position of signal peptide, transmembrane region in TM, and the furin cleavage site between SU and TM subunits were determined by dedicated prediction servers [12-14]. SU, surface glycoprotein; TM, transmembrane glycoprotein; polyA, polyadenylation site.

Additional file 2: ELVgv consensus sequence. Sequence identical to Additional file 1, provided in simple text format.

Additional file 3: (Upper) RNA secondary structure motifs predicted by mfold thermodynamic folding algorithm, with the associated change in Gibbs free energy (dG) [10]. (Lower) Alignment of homologous regions of the deduced amino acid sequence of ELVgv *orf2* with HIV-1 Tat [GenBank:ABF00629.1].

Additional file 4: Alignment of retroviral RT amino acid sequences. The alignment of 16 retroviral RT amino acid sequences used to generate the tree in Figure 3A. RT domains from the following viruses were used: HIV-1 [GenBank:K03455]; simian immunodeficiency virus, SIV [GenBank:NC758887]; FIV [GenBank:NC001482]; equine infectious anemia virus, EIAV [GenBank:NP_056902]; Visna/maedi virus [GenBank:NC001452]; caprine arthritis encephalitis virus, CAEV [GenBank:NP_040939]; bovine immunodeficiency virus, BIV [GenBank:NP_040563]; Rous sarcoma virus, RSV [GenBank:AF033808]; Mason-Pfizer monkey virus, MPMV[GenBank:NC001550]; murine leukemia virus, MLV [GenBank:NC001501]; walleye dermal sarcoma virus, WDSV [GenBank:AF033822]; simian foamy virus, SFV [GenBank:NC001364]; bovine leukemia virus, BLV [GenBank:NC_001414]; human T-lyphotropic virus 1, HTLV-1 [GenBank:D13784]; HTLV-2 [GenBank:M10060].

Additional file 5: Alignment of lentiviral gag-pol nucelotide sequences. The alignment of 35 conserved regions of *gag* and *pol* lentiviral sequences from Gilbert *et al.* [5], together with the recently described ELVmpf [6,8] and the ELVgv described here. This alignment was used to generate the tree in Figure 3B.

Additional file 6: Phylogenetic relationship of ELVgv to other lentiviruses using Bayesian analysis. The same alignment of 33 lentiviral *gag-pol* sequences as in ML analysis in Figure 3B was used. Bayesian analysis was run for 1,000,000 steps, sampling every 5,000 steps and discarding first 25% of the trees. Average standard deviation of split frequencies converged bellow 0.001. The GTR+I+gamma nucleotide model (a General Time Reversible model with a proportion of invariable sites and a gamma-shaped distribution of rates across sites) was employed in MrBayes program [18]. Appropriate model was selected using program jModeltest vs (Darriba D, et al: jModelTest2: more models, new heuristics and parallel computing. Nat Methods 2012, 9:772). Values of posterior probabilities are shown. Grey lines designate groups of exogenous lentiviruses. The endogenous lentiviruses are denoted by black dots. Scale bar indicates number of nucleotide substitutions per site.

Additional file 7: Phylogenetic relationship of three full-length ELVgv insertions to other lentiviruses using ML analysis. The same alignment of lentiviral *gag-pol* sequences as in ML analysis in Figure 3B was used, with the ELVgv consensus sequence substituted by sequences of proviruses *I*, *II* and *III*. The alignment was generated in MEGA5 program [15] using the MUSCLE algorithm [16]. The ML analysis was performed using MEGA5 program under Tamura-Nei model, Nearest-Neighbor-Interchange ML heuristic method and otherwise default parameters. Bootstrap supports (percent out of 1,000 replicates) are shown. Grey lines designate groups of exogenous lentiviruses. The endogenous lentiviruses are denoted by black dots. Scale bars indicate number of nucleotide substitutions per site.

Additional file 8: List of ELVgv insertions residing in regions of putative genome duplication. The table lists accession numbers for

each pair of colugo genomic contigs that show large regions of apparent segmental duplications (size up to ~ 20 kb). All contigs harbor ELVgv solo LTR sequences in the duplicated regions. The genetic distance of the duplicated region was used to estimate the age of the duplication event, using the same formula as for the estimates based on LTR sequences.

Competing interests
The authors declare that they have no competing interest.

Authors' contributions
DE and JP designed the study. All authors participated in the data collection and analysis, and in writing of the manuscript. All authors read and approved the final manuscript.

Acknowledgements
We would like to acknowledge Richard K. Wilson and The Genome Institute, Washington University School of Medicine, for the generation and public release of the Galeopterus sequence assembly. We thank to Jiří Hejnar and members of his laboratory for helpful comments to the manuscript. This work was funded by program LK11215 provided by the Czech Ministry of Education, Youth and Sports. Access to computing and storage facilities provided by ELIXIR CZ and the National Grid Infrastructure MetaCentrum, administered under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005), is greatly appreciated.

Received: 30 June 2014 Accepted: 9 September 2014
Published online: 04 October 2014

References
1. Johnson WE: **A proviral puzzle with a prosimian twist.** *Proc Natl Acad Sci U S A* 2008, **105**(51):20051–20052.
2. Stoye JP: **Studies of endogenous retroviruses reveal a continuing evolutionary saga.** *Nat Rev Microbiol* 2012, **10**(6):395–406.
3. Katzourakis A, Tristem M, Pybus OG, Gifford RJ: **Discovery and analysis of the first endogenous lentivirus.** *Proc Natl Acad Sci U S A* 2007, **104**(15):6261–6265.
4. Gifford RJ, Katzourakis A, Tristem M, Pybus OG, Winters M, Shafer RW: **A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution.** *Proc Natl Acad Sci U S A* 2008, **105**(51):20362–20367.
5. Gilbert C, Maxfield DG, Goodman SM, Feschotte C: **Parallel germline infiltration of a lentivirus in two Malagasy lemurs.** *PLoS Genet* 2009, **5**(3):e1000425.
6. Cui J, Holmes EC: **Endogenous lentiviruses in the ferret genome.** *J Virol* 2012, **86**(6):3383–3385.
7. Keckesova Z, Ylinen LM, Towers GJ, Gifford RJ, 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* 2009, **384**(1):7–11.
8. Han GZ, Worobey M: **Endogenous lentiviral elements in the weasel family (Mustelidae).** *Mol Biol Evol* 2012, **29**(10):2905–2908.
9. Belshaw R, Watson J, Katzourakis A, Howe A, Woolven-Allen J, Burt A, Tristem M: **Rate of recombinational deletion among human endogenous retroviruses.** *J Virol* 2007, **81**(17):9437–9442.
10. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Res* 2003, **31**(13):3406–3415.
11. Kohany O, Gentles AJ, Hankus L, Jurka J: **Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor.** *BMC Bioinformatics* 2006, **7**:474.
12. Petersen TN, Brunak S, Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nat Methods* 2011, **8**:785–786.
13. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *J Mol Biol* 2001, **305**(3):567–580.
14. Duckert P, Brunak S, Blom N: **Prediction of proprotein convertase cleavage sites.** *Protein Eng Des Sel* 2004, **17**(1):107–112.
15. 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.** *Mol Biol Evol* 2011, **28**(10):2731–2739.
16. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**(5):1792–1797.
17. Dimmic MW, Rest JS, Mindell DP, Goldstein RA: **rtREV: an amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny.** *J Mol Evol* 2002, **55**(1):65–73.
18. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics (Oxford, England)* 2001, **17**(8):754–755.
19. Johnson WE, Coffin JM: **Constructing primate phylogenies from ancient retrovirus sequences.** *Proc Natl Acad Sci U S A* 1999, **96**(18):10254–10260.
20. Kumar S, Subramanian S: **Mutation rates in mammalian genomes.** *Proc Natl Acad Sci U S A* 2002, **99**(2):803–808.
21. Chinwalla AT, Cook LL, Delehaunty KD, Fewell GA, Fulton LA, Fulton RS, Graves TA, Hillier LW, Mardis ER, McPherson JD, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, *et al*: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**(6915):520–562.
22. Janečka JE, Helgen KM, Lim NT-L, Baba M, Izawa M, Boeadi, Murphy WJ: **Evidence for multiple species of Sunda colugo.** *Curr Biol* 2008, **18**(21):R1001–R1002.
23. Martin RD: **Colugos: obscure mammals glide into the evolutionary limelight.** *J Biol* 2008, **7**(4):13.
24. Nie WH, Fu BY, O'Brien PCM, Wang JH, Su WT, Tanomtong A, Volobouev V, Ferguson-Smith MA, Yang FT: **Flying lemurs - the 'flying tree shrews'? molecular cytogenetic evidence for a Scandentia-Dermoptera sister clade.** *BMC Biol* 2008, **6**:11.
25. Janecka JE, Miller W, Pringle TH, Wiens F, Zitzmann A, Helgen KM, Springer MS, Murphy WJ: **Molecular and genomic data identify the closest living relative of primates.** *Science (New York, NY)* 2007, **318**(5851):792–794.
26. Hedges SB, Dudley J, Kumar S: **TimeTree: a public knowledge-base of divergence times among organisms.** *Bioinformatics (Oxford, England)* 2006, **22**(23):2971–2972.

doi:10.1186/s12977-014-0084-x
Cite this article as: Hron *et al.*: Endogenous lentivirus in Malayan colugo (*Galeopterus variegatus*), a close relative of primates. *Retrovirology* 2014 **11**:84.

Life History of the Oldest Lentivirus: Characterization of ELVgv Integrations in the Dermopteran Genome

Tomáš Hron[1], Helena Farkašová[1], Abinash Padhi[2], Jan Pačes[1], and Daniel Elleder*[1]

[1]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic

[2]Department of Animal and Avian Sciences, University of Maryland, College Park

***Corresponding author:** E-mail: daniel.elleder@img.cas.cz.

Associate editor: Harmit Malik

Abstract

Endogenous retroviruses are genomic elements formed by germline infiltration by originally exogenous viruses. These molecular fossils provide valuable information about the evolution of the retroviral family. Lentiviruses are an extensively studied genus of retroviruses infecting a broad range of mammals. Despite a wealth of information on their modern evolution, little is known about their origins. This is partially due to the scarcity of their endogenous forms. Recently, an endogenous lentivirus, ELVgv, was discovered in the genome of the Malayan colugo (order *Dermoptera*). This represents the oldest lentiviral evidence available and promises to lead to further insights into the history of this genus. In this study, we analyzed ELVgv integrations at several genomic locations in four distinct colugo specimens covering all the extant dermopteran species. We confirmed ELVgv integrations in all the specimens examined, which implies that the virus originated before the dermopteran diversification. Using a locus-specific dermopteran substitution rate, we estimated that the proviral integrations occurred 21–40 Ma. Using phylogenetic analysis, we estimated that ELVgv invaded an ancestor of today's *Dermoptera* in an even more distant past. We also provide evidence of selective pressure on the *TRIM5* antiviral restriction factor, something usually taken as indirect evidence of past retroviral infections. Interestingly, we show that *TRIM5* was under strong positive selection pressure only in the common dermopteran ancestor, where the ELVgv endogenization occurred. Further experiments are required to determine whether ELVgv participated in the *TRIM5* selection.

Key words: **endogenous retrovirus, dermoptera, TRIM5.**

Introduction

Endogenous retroviruses (ERVs) are genomic elements present in all the vertebrate genomes hitherto analyzed. They are formed when an originally exogenous retrovirus integrates into the host germline genome and becomes transmitted from generation to generation. The vast majority of these "viral fossils" are of ancient origin and thus provide valuable insight into the evolutionary history of retroviruses and their interactions with hosts. The phylogenetic structure of ERVs matches that of the main genera of the exogenous retroviruses hitherto described. However, the diversity of ERVs is much greater, as they represent the footprints of viral infections that occurred throughout the entire retroviral evolution (Hayward et al. 2015; Johnson 2015). The basic approach to uncovering the origin and phylogenetic relationships between particular retroviral genera is through the genetic analysis of ERV lineages.

There are several methods to estimate the time interval when a specific ERV infiltrated the germline of its host species. The most straightforward way is to determine the presence or absence of ERV in the genomes of phylogenetically related species (Johnson 2015). In general, any ERV infiltration should have occurred in the most recent common ancestor of all the ERV-positive species. This method is very robust, but it only yields a broad interval for the ERV age. A second approach, which yields age estimates for specific ERV integrations, involves the sequence analysis of viral long terminal repeats (LTRs). The 5′ and 3′ LTRs are formed during retroviral reverse transcription and should be identical at the time of integration. Integration can therefore be dated using the LTRs' sequence divergence generated post-integration at the genome substitution rate (Johnson and Coffin 1999). A third method uses the fact that ERV integration can be dated if the genomic locus with the ERV integration is duplicated. If the virus is present in both regions, then the duplication event must have occurred after virus integration. Therefore, the age of the duplication provides a lower estimate for the ERV integration (Katzourakis et al. 2007; Hron et al. 2014). The fourth—and the most sophisticated—approach involves the time-calibrated phylogenetic analysis of orthologous proviral sequences from multiple species (Tonjes and Niebert 2003; Jha et al. 2009, 2011; Kamath et al. 2014). In addition to describing individual integrations, this method is able to reveal the evolutionary history of the entire ERV lineage.

Apart from the study of ERV sequences, additional indirect evidence for ancient retroviral infections can be obtained by identifying positively selected host antiretroviral factors (Aswad and Katzourakis 2012). This is based on the

Article

assumption that antiviral factors are forced to continually adapt to the infecting virus. A great example of such virus-host coevolution is provided by the TRIM5 restriction factor, which blocks retroviral infection by binding the capsid (CA) component of the viral core (Stremlau et al. 2004; Perez-Caballero et al. 2005). The antiretroviral activity of this protein is determined mainly by the B30.2 domain, which has been clearly shown to have been subjected to positive selection pressure (Sawyer et al. 2005). In fact, *TRIM5* evolution also reflects ancient lentiviral infections. In lagomorphs (rabbits and hares) and primates, the pattern of positive selection and/or antiretroviral activity of TRIM5 have been linked to the presence of endogenous lentiviruses (Fletcher et al. 2010; de Matos et al. 2011; Rahm et al. 2011; Yap and Stoye 2013; McCarthy et al. 2015).

Lentiviruses represent one of the most extensively studied retroviral genera, mainly because of their ability to cause various chronic diseases in a broad range of mammals. They also constitute a major public health concern due to the HIV/AIDS pandemic. Until recently, lentiviruses were regarded as a young retroviral genus and no endogenous counterparts were known. This changed in 2007, when the first endogenous lentivirus was discovered in the rabbit genome (Katzourakis et al. 2007). There followed the discoveries of other endogenous lentiviral lineages in lemurs, ferrets and colugos (Gifford et al. 2008; Cui and Holmes 2012; Hron et al. 2014). All these lineages are quite distant from the extant lentiviruses, forming deep, separate branches in the lentiviral phylogeny.

The lemur virus, pSIV, is the only documented case of endogenous lentivirus in primates. Its individual proviral sequences integrated ~3–6 Ma (Gifford et al. 2008; Gilbert et al. 2009). Lentiviral endogenization in rabbits and ferrets seems to have occurred deeper in the past—~12 Ma (Keckesova et al. 2009; van der Loo et al. 2009; Cui and Holmes 2012; Han and Worobey 2012). The oldest known lentiviral lineage is represented by the colugo endogenous lentivirus, ELVgv. The dates of the individual ELVgv integrations fall into the 5–30 Ma interval (Hron et al. 2014; Han and Worobey 2015). In all of the earlier-mentioned cases, no detailed phylogenetic analysis was performed that could be used to derive the evolutionary history of the ERV lineage.

The Malayan colugo (*Galeopterus variegatus*), the modern descendant of an ancient ELVgv host, is a mammal found in Southeast Asia. It belongs to the order *Dermoptera*, which contains only two extant genera, *Galeopterus* and *Cynocephalus*, which split ~15 Ma (Janecka et al. 2008). The *Dermoptera* are considered the closest order to the order primates, from which they split ~81 Ma (Hedges et al. 2006; Janecka et al. 2007).

In this work, we focus on the endogenous lentivirus ELVgv in the colugo genome. It is the oldest lentiviral lineage described to date and can therefore provide a valuable insight into the origins of lentiviruses. Using genomic DNA samples from multiple dermopteran species, we uncovered the evolutionary history of ELVgv and estimated the time when it invaded the host genome. To obtain indirect evidence of past retroviral infections in Dermoptera, we also sequenced the dermopteran *TRIM5* restriction factor and described the signature of its positive selection.

Results and Discussion

Detection of ELVgv Integrations in *Dermoptera*

In our previous work, we showed the presence of ELVgv in the genomic draft assembly of Malayan colugo (Hron et al. 2014). This is the only dermopteran species whose genome assembly is publicly available (GenBank: GCA_000696425; the specimen that was sequenced is designated here as GVAgb). To narrow down the time interval of ELVgv endogenization, we decided to detect its sequence in the genomes of different colugo specimens. Three samples of genomic DNAs, covering both of the extant dermopteran genera, were kindly provided by W. Murphy (Texas A&M University). These included two *G. variegatus* subspecies (designated as GVA3, and GVA5) and *Cynocephalus volans* (CVO) (Janecka et al. 2008). Due to the very low amounts of DNA available, we performed most of the following experiments with whole genome amplified samples (see Materials and Methods section).

First, we checked for the presence of an ELVgv proviral sequence in the dermopteran samples. This was done by PCR amplification with two primer pairs targeting the ELVgv reverse transcriptase (RT), the most conserved retrovirus gene. Interestingly, we confirmed the presence of ELVgv sequences in all the genomic samples analyzed (fig. 1; RT1 and RT2). To describe the pattern of particular proviral integrations in *Dermoptera*, we then focused on three full-length ELVgv copies previously described in GVAgb genome assembly (designated here as ***A***, ***B*** and ***C***, corresponding to proviruses ***I***, ***II*** and ***III***, respectively, in the previous study) (Hron et al. 2014). Using primers that specifically amplify the junctions of viral 3′-LTR and dermopteran genomic DNA, we detected all three proviral integrations in each dermopteran specimen (fig. 1; 3′ junction). PCR analysis with primers detecting empty pre-integration sites did not yield any product, which implies that all three integrations are homozygous in the animals tested. To exclude cross-contamination and wrong identification of genomic DNA in our samples, we PCR-amplified and sequenced control genomic loci in dermopteran genes previously described (Janecka et al. 2008). Their comparison with previously published sequences for individual species was in good agreement (supplementary fig. S1, Supplementary Material online), confirming the identity of the genomic samples and the absence of contamination.

These results imply that the ***A***, ***B*** and ***C*** integrations all occurred in the common ancestor of *Galeopterus* and *Cynocephalus* and therefore must be older than the split of both species (~15.3 Ma; Hedges et al. 2006). These are surprising findings, considering the fact that the provirus ***A*** integration was previously estimated to have taken place in the interval of 5.1–10.3 Ma (Hron et al. 2014).

Sequencing of ELVgv Proviruses in Dermopteran Specimens

To obtain more information for each of the ELVgv integrations (***A***, ***B*** and ***C***), we decided to sequence proviral DNA

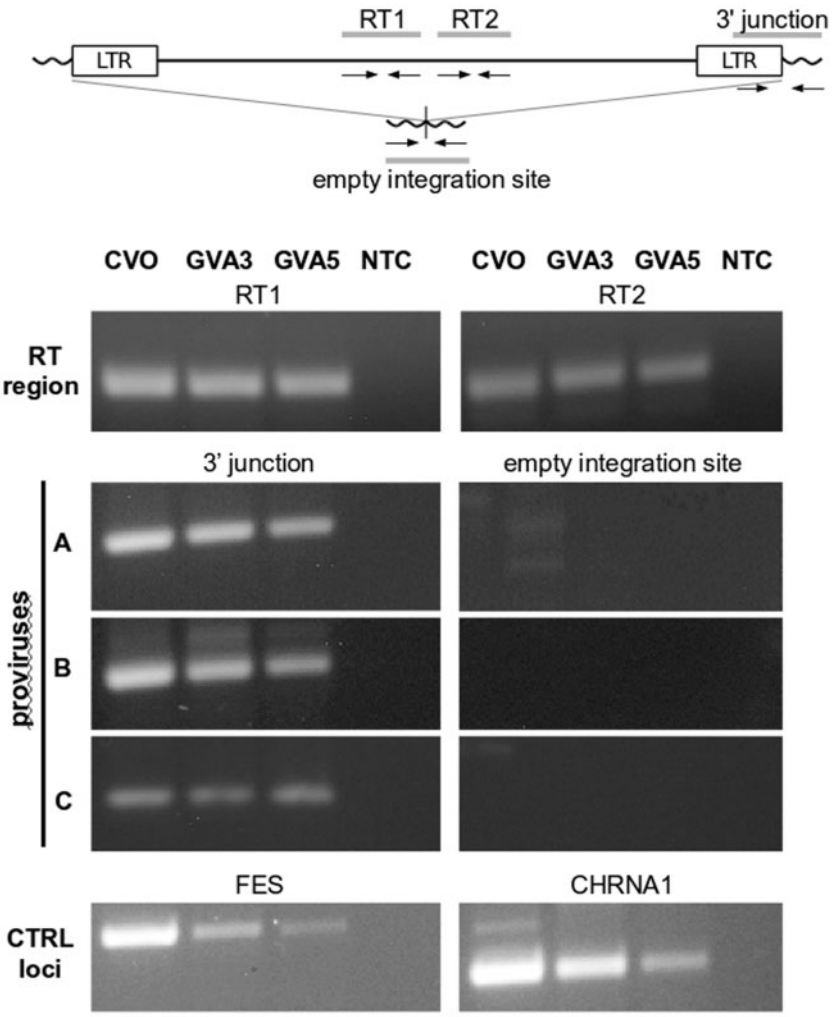

Fig. 1. Detection of ELVgv integrations in the dermopteran genomic samples. *Upper*: schematics of ELVgv provirus with primer positions. *Lower*: PCR amplifications using primers targeting two regions of the ELVgv reverse transcriptase (RT), the three individual virus-host junctions and their corresponding pre-integration sites (proviruses *A*, *B*, and *C*), and two control loci (*FES*, *CHRNA1*) in dermopteran genome (Janecka et al. 2008). CVO, *C. volans* specimen; GVA3, GVA5, *G. variegatus* specimens; NTC, nontemplate PCR control.

from all three dermopteran individuals (GVA3, GVA5, CVO). Due to the limited amount and fragmentary nature of whole genome-amplified DNA, we decided to sequence only the 5′ parts of the proviruses (fig. 2). Using a set of primers anchored in the 5′ flank of the dermopteran genome and in the proviral regions, we amplified and sequenced four ELVgv fragments ~3 kilobases (kb) in length and four shorter fragments 0.7–1 kb in length. In the GVA3 specimen, we were able to amplify only the very end of the provirus ***C*** sequence, probably either due to the low quality of whole genome-amplified template DNA or due to the mutations/deletions in the regions targeted by the primers. The short sequence amplified was excluded from further analyses. All together, we obtained a total of 11 partial ELVgv sequences from four animals, including three provirus sequences reconstructed in silico in GVAgb genome assembly (GenBank: KX022581-9).

The ***A***, ***B*** and ***C*** proviruses, which represent independent integration events, are mutually different, as described before (Hron et al. 2014). All proviral sequences analyzed contain number of nonsense mutations and indels in their coding regions (supplementary fig. S2, Supplementary Material online). Moreover, each provirus differs in many sites between individual animals (fig. 2). These differences have accumulated in the provirus after the split of the dermopteran lineages analyzed. The differences detected include: substitutions, short indels and three long insertions (formed by integrations of SINE and LINE repetitive elements). For each provirus, the sequences in CVO always substantially differ from the sequences in *Galeopterus* (GVA3, GVA5, GVAgb), reflecting the separate evolution of the *Cynocephalus* and *Galeopterus* species. The pattern of sequence differences observed for each proviral integration enabled us to analyze the ELVgv evolutionary history in *Dermoptera*.

Improved Timing Estimates of ELVgv Integrations

As mentioned earlier, the fact that the ***A***, ***B***, and ***C*** proviruses can be detected in both the *Cynocephalus* and *Galeopterus* genomic DNA samples indicates that they all entered the

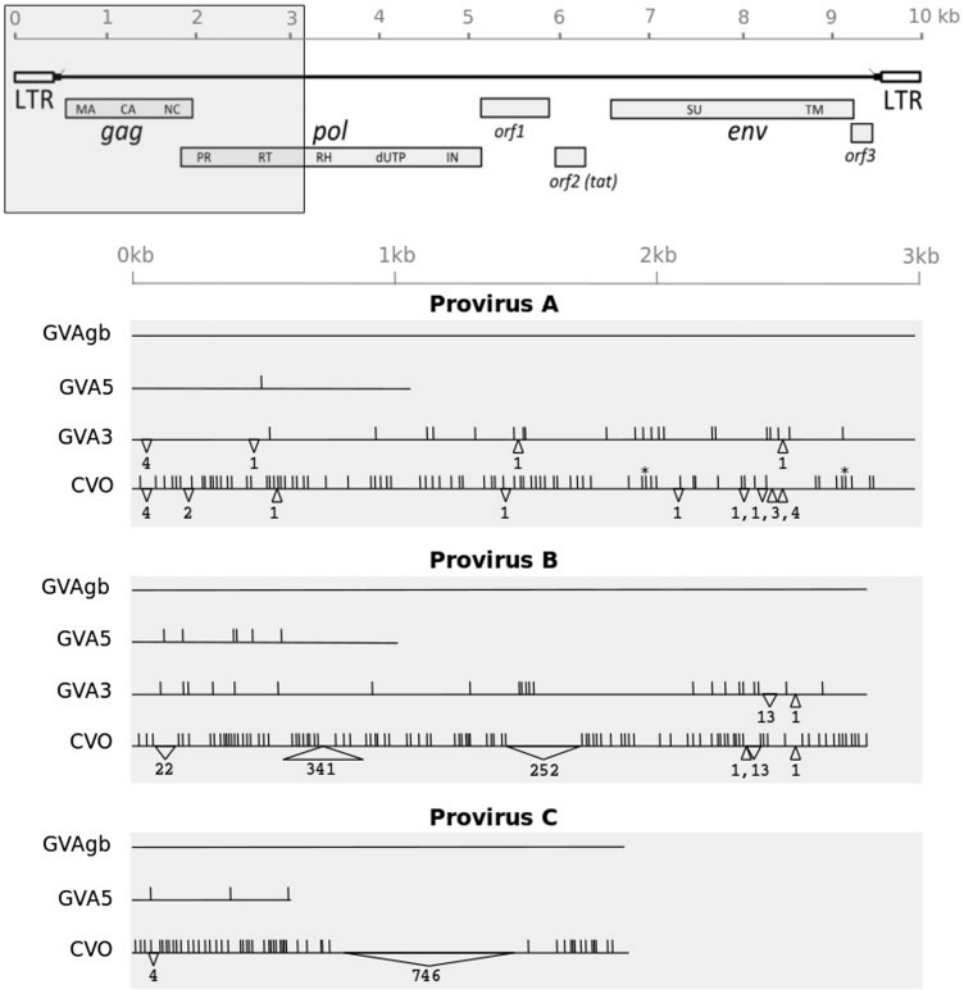

FIG. 2. Sequence variability of three ELVgv proviruses integrated in the dermopteran genome. *Upper*: Schematic depiction of ELVgv genomic organization (Hron et al. 2014). Region analyzed by sequencing is highlighted by grey box. *Lower*: Graphical representation of individual partial proviral sequences, with scale indicated earlier. Each line represents one proviral sequence in particular specimen (CVO, GVA3, GVA5, GVAgb). Three blocks of lines correspond to the three proviral integrations (*A*, *B*, and *C*), where the sequence from GVAgb is taken in each case as a reference. Vertical lines indicate single nucleotide substitutions relative to the reference sequence. Insertions and deletions are depicted by upward and downward-pointing triangles, respectively, with the length indicated below. Heterozygous sites are indicated by asterisks.

dermopteran germline before the two genera diverged (~15.3 Ma). Furthermore, ELVgv is limited exclusively to *Dermoptera* and is not found in other mammals (Hron et al. 2014). Since *Dermoptera* split from the other mammals ~81.3 Ma (Hedges et al. 2006), the time interval when ELVgv endogenization must have occurred is 15.3–81.3 Ma.

Previously, the LTR-aging method has been used to date ELVgv integrations (Hron et al. 2014). The method used a range of average mammalian substitution rates ($2.2–4.5\times10^{-9}$ per site per year). However, this range represents only a rough estimate, as substitution rates greatly vary across species and genomic loci (Hodgkinson and Eyre-Walker 2011). We therefore decided to calculate relevant locus-specific substitution rates to obtain more accurate dates for ELVgv integrations (Martins and Villesen 2011). As mentioned earlier, we partially sequenced ELVgv proviruses in both *Galeopterus* and *Cynocephalus*, whose divergence time is known. This enabled us to estimate the substitution rate for each proviral locus (table 1; details available in Materials and Methods). These estimated substitution rates were 0.95, 1.25 and 1.57×10^{-9} per site per year for provirus ***A***, ***B*** and ***C***, respectively. For control purposes, we also calculated substitution rates for several dermopteran genes for which both *Galeopterus* and *Cynocephalus* sequences are publicly available, obtaining values in the range of $0.732–1.34\times10^{-9}$ per site per year. All the estimates we obtained show only slight variation and are substantially lower than the average mammalian rates used in previous works. These new low values are consistent with a recent study that proposed an average dermopteran substitution rate of 0.69×10^{-9} per site per year (Perelman et al. 2011).

We used the LTR-aging method with these new values and obtained more accurate estimates for the ***A***, ***B*** and ***C*** integration ages: 21.1, 37.6 and 40.0 Ma, respectively (table 1). This almost doubles the ages suggested previously (Hron et al. 2014). This is consistent with the earlier-mentioned broad time interval for ELVgv endogenization (15.3–81.3 Ma). On the other hand, the LTR-aging method may yield too low an

Table 1. Dermopteran Substitution Rate Calculation and LTR Aging of ELVgv Proviruses.

Provirus	Divergence between GVA and CVO [substitutions/nt]				Substitution Rate [substitutions/nt/year]	Integration Time Estimation Based on LTR Aging	
	GVAgb/CVO	GVA5/CVO	GVA3/CVO	Average		LTR Divergence [substitution/nt]	Age Estimate [millions of years]
A	0.0263	0.0328	0.0281	0.0291	0.95×10^{-9}	0.0401	21.1
B	0.0382	0.0384	0.0379	0.0381	1.25×10^{-9}	0.0937	37.6
C	0.0462	0.0500	–	0.0481	1.57×10^{-9}	0.1259	40.0

estimate in cases where gene conversion between 5′ and 3′ LTRs occurred (Kijima and Innan 2010). Moreover, the estimated ages of very old integrations may be imprecise, as the more recent substitution rates on which they are based might not apply.

More information about the age of ERV integration can be obtained in cases where the genomic locus containing ERV integration site is duplicated. This has previously been used to corroborate the age estimates for RELIK infiltration into the rabbit genome (Katzourakis et al. 2007) and for ELVgv infiltration into the colugo genome (Hron et al. 2014). The weak point of this method is that it is dependent on the correct genomic assembly of the duplicated regions, a task made extremely difficult due to their high degree of similarity.

In view of this, we proceeded to experimentally confirm the presence of the putative ELVgv-containing duplications in our genomic DNA dermopteran samples using quantitative digital droplet PCR (ddPCR) method. Out of the seven previously described ELVgv integrations in duplicated regions (Hron et al. 2014), only two were suitable for PCR primer design due to their extremely high genomic repeat content. One of these regions (region G2; supplementary fig. S3, Supplementary Material online) shows copy number values consistent with duplication or even multiplication. However, the other region (region G1; supplementary fig. S3, Supplementary Material online) has a copy number consistent with single copy genomic locus. For control purposes, we also tested two out of three putatively duplicated RELIK integrations in the rabbit genome (Katzourakis et al. 2007). Both of these regions yielded copy number values confirming their multiplication. These data indicate that duplications implied by genomic assemblies might be incorrect. Their confirmation by experiment is, therefore, required for this aging method to be reliable.

Evolutionary History of ELVgv

It is generally assumed that all proviruses forming a single ERV group are the descendants of a single original integration (Jern and Coffin 2008). To describe the mutual relationships of ***A***, ***B*** and ***C*** proviruses, we constructed their maximum likelihood (ML) phylogeny (fig. 3*B*). The tree topology was strongly supported by high bootstrap values. As expected, the branches corresponding to each provirus cluster monophyletically, which is consistent with the finding that their integrations precede the split of the species analyzed. The tree also indicates that the most recent common ancestor (MRCA) of ***A*** and ***B*** originated from the MRCA of all three proviruses. Additionally, divergences between 5′ and 3′ LTRs (table 1), which reflect the accumulated evolutionary changes following proviral integration, were used to identify points corresponding to the provirus integrations. The distance was measured from the tips of the GVAgb branches, for which the LTR aging had been calculated (fig. 3*B*; arrows). As expected, the estimated integrations lie, in all three cases, on the branch corresponding exclusively to the provirus concerned and precede the MRCA of the species analyzed. The ML phylogeny was further corroborated using the Bayesian approach. The tree topology and branch lengths of the Bayesian tree were in good agreement with the ML tree (data not shown).

According to the molecular clock hypothesis, the nodes of a phylogenetic tree can be calibrated using known evolutionary events, such as species divergence times (Drummond et al. 2006). Our ML phylogeny, coupled with the use of the *Galeopterus*/*Cynocephalus* divergence time as a calibration constraint, allowed us to estimate the time when different ELVgv lineages originated (fig. 3*A*). The ages of the ***A***/***B*** and ***A***/***B***/***C*** MRCAs were estimated at ∼40 and 60 Ma, respectively. We corroborated the ML analysis by the Bayesian approach using BEAST2 (Bouckaert et al. 2014). This approach yielded the estimated ages of 37.1 Ma (95% highest posterior density interval 27.6–46.0) for the ***A***/***B*** MRCA and 53.1 Ma (38.1–65.9) for the ***A***/***B***/***C*** MRCA. These intervals are consistent with the ML estimates.

Taken together, this phylogenetic analysis allowed us to estimate the evolutionary ages of individual ELVgv integrations and the relationships between them. We show that ELVgv infiltration of the dermopteran germline occurred approximately between 60 and 20 Ma. However, these data should be regarded with caution. One reason is that the virus life history from the MRCA until the time of the provirus integration may have involved exogenous stages, i.e., the endogenous copy may have undergone several cycles of reinfection or retrotransposition. During such stages, the virus would have been subjected to exogenous retrovirus substitution rates, which are six orders of magnitude higher than neutral host substitution rates (Li et al. 1988). This would have reduced the estimated MRCA-to-integration interval. This uncertain part of ELVgv's life history is denoted by the dashed line (fig. 3*A*).

The time we propose for ELVgv endogenization implies that this lentiviral lineage infected the ancestors of the extant *Dermoptera* after their split from the other mammalian orders. As ELVgv is the most ancient lentivirus identified to date, the genus *Lentivirus* originated >40 Ma, possibly even >60 Ma, i.e., before the beginning of the Cenozoic era.

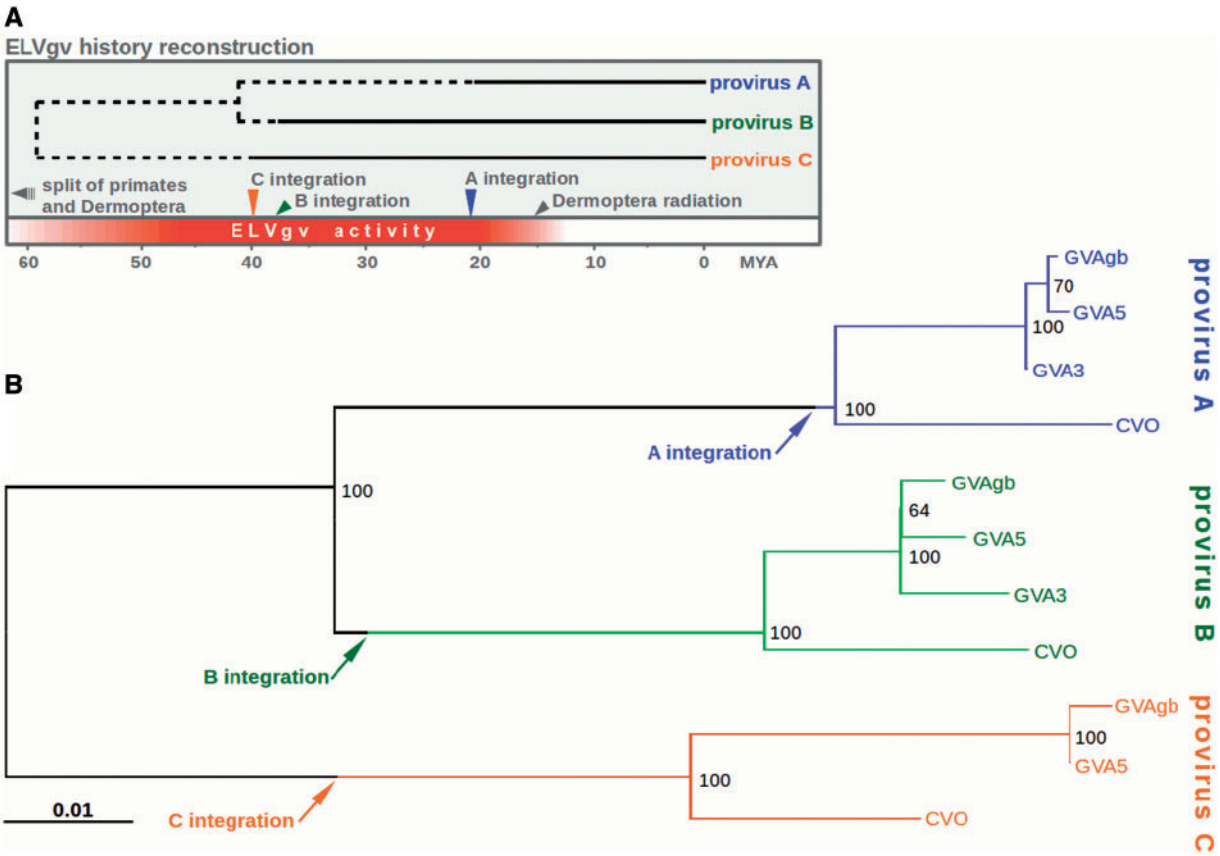

FIG. 3. Evolutionary analysis of three ELVgv integrations. (*A*) Schematic depiction of ELVgv life history. The bottom axis depicts evolutionary time since present in millions of years. Ages of the most recent common ancestors of analyzed proviruses were estimated from ML phylogeny by assuming a known time of the *Galeopterus*/*Cynocephalus* split. Dashed line indicates the time preceding the integration of each provirus. Various evolutionary events are shown in the bottom part of the scheme. (*B*) ML tree of ELVgv proviral nucleotide sequences. Different colors correspond to the three proviral integrations (*A*, *B*, and *C*). Numbers next to nodes correspond to bootstrap values. Points corresponding to the estimated integration events of each provirus (details in main text) are depicted by arrows. Scale bar represents number of substitutions per nucleotide.

This age is still much lower than the presumed origin of the rest of the retroviral family, which is estimated to be >100 My old (Katzourakis et al. 2009; Johnson 2015). If more endogenous lentiviral sequences are found, it is possible that the origin of the lentiviral genus will be moved even further into the past.

Identification and Positive Selection Analysis of Dermopteran *TRIM5*

As the long-term occurrence of retroviral infections usually creates a selective pressure on host antiretroviral factors, we tried to investigate the pattern of positive selection acting on one of the most studied host restriction factors, TRIM5. We chose the longest TRIM5α protein isoform, which was reported to have the restriction activity (Stremlau et al. 2004). First, we assembled a *TRIM5α* coding sequence (GenBank: BK009414) from publicly available RNA-seq data from GVAgb (details in Materials and Methods). This sequence is predicted to encode a protein of 486 amino acids, with one noncoding and seven coding exons, similarly to human *TRIM5α* structure. Using this sequence information, we designed primers to amplify and sequence *TRIM5α* coding regions from the dermopteran genomic DNA samples (GVA3, GVA5, CVO). In addition to the *TRIM5α* sequence from GVAgb reconstructed in silico, we obtained >91% of the *TRIM5α* coding sequences from GVA3, GVA5 and CVO (GenBank: KX022578-80). To check that we sequenced real orthologs of *TRIM5* and not some other member of the TRIM family, we constructed a ML tree for the dermopteran sequences along with *TRIM5* and its closest paralogs (*TRIM22* and *TRIM33*) from other mammals (supplementary fig. S4, Supplementary Material online). This constitutes the first experimental identification of an antiviral restriction factor from the order *Dermoptera*.

To analyze the selection pressures acting on the TRIM5 restriction factor, we used the dataset of dermopteran coding sequences together with sequences from selected primates, which represent the closest order to the order *Dermoptera*. The ML tree generated from these *TRIM5α* sequences showed the well-supported topology corresponding to the phylogenetic relationships expected between the species (fig. 4).

The general measure of natural selection is the dN/dS, i.e., the ratio of the nonsynonymous substitution rate (dN) and the synonymous substitution rate (dS) pertaining to the coding sequence analyzed. To estimate dN/dS values for our dataset, we used the PAML 4.7 software (Yang 2007) and the Datamonkey web server (Pond and Frost 2005a; Delport et al. 2010). In agreement with previous studies

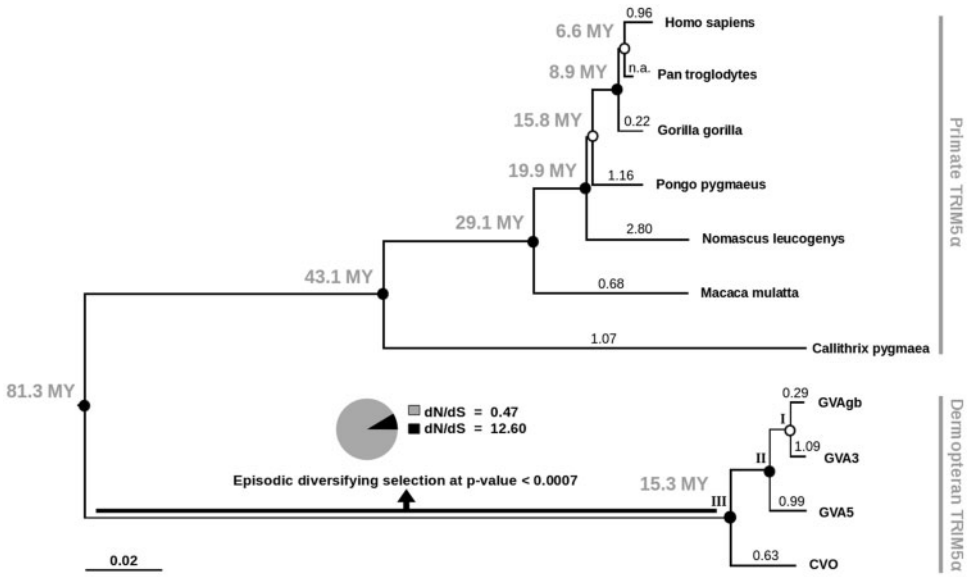

Fig. 4. Analysis of selective pressures acting on *TRIM5α* sequence in dermopteran and primate species. ML tree of *TRIM5α* coding nucleotide sequences is shown. Bootstrap support values are depicted either as closed circles (bootstrap support > 95%) or open circles (bootstrap < 95%) for each node. Ages of most recent common ancestors, suggested in TimeTree (Hedges et al. 2006), are denoted next to the corresponding nodes in millions of years (MY). Values above terminal branches correspond to mean dN/dS estimates calculated in Datamonkey. This values do not differ by > 10% from the results obtained using PAML (data not shown). Result from the branch-site REL test for episodic positive selection performed in Datamonkey is shown for the branch leading to the dermopteran clade (node III). Specifically, dN/dS estimates of the two classes of codons with the percent of codons falling into each class are depicted. Scale bar represents number of substitutions per nucleotide.

Table 2. Significance of Positive Selection Pressure Acting on *TRIM5α* in Particular Primate and Dermopteran Lineages.

Clade	Test of Positive Selection (PAML)[a]			Test of Episodic Positive Selection (Datamonkey)[b]		
	dN/dS (%)[c]	LRT	*P* value[d]	dN/dS (%)[c]	LRT	*P* value[d]
GVAgb	n.a.	0.000	1.0000	737.2 (0.3%)	7.988	0.0400*
GVA3	n.a.	0.000	1.0000	1.8 (43.0%)	0.068	1.0000
GVA5	n.a.	0.000	1.0000	1.1 (80.1%)	0.014	1.0000
CVO	n.a.	0.000	1.0000	275.2 (0.4%)	1.868	0.8587
Node I	8.0 (11.1%)	0.630	0.4275	3.0 (66.8%)	0.677	1.0000
Node II	n.a.	0.000	1.0000	25.8 (0.9%)	1.979	0.9567
Node III	19.6 (5.1%)	10.352	0.0013**	12.6 (10.4%)	15.686	0.0007***
Dermoptera	5.8 (2.6%)	2.233	0.1351	–	–	–
Primates	4.0 (16.3%)	23.402	<0.0001****	–	–	–

[a]Branch-site test of positive selection in codeml program of PAML package.
[b]Branch-site REL test for episodic positive selection in Datamonkey web server.
[c]dN/dS ratio estimate of the class of codons under positive selection with the percentage of codons falling into this class designated in parentheses.
[d]*P* values calculated from likelihood ratio test (LRT) statistics, level of significance is expressed by asterisk.

(Sawyer et al. 2007), the overall dN/dS calculated for each branch (fig. 4) showed alternating patterns of positive selection among primate species. In the dermopteran branch, the values do not point to strong positive selection.

In general, the majority of sites in a protein are subjected to purifying selection pressure conserving the protein function and only a few sites are free to diversify (Yang et al. 2000). Therefore, we also employed more sensitive branch-site model analysis assuming the variation of dN/dS in each sequence (Yang and Nielsen 2002; Zhang et al. 2005) and calculated the significance of positive selection for each branch using likelihood-ratio test. Specifically, we used the branch-site model of positive selection (Zhang et al. 2005) in the PAML software and the branch-site Random Effect Likelihood (REL) test for episodic positive selection (Pond et al. 2011) in Datamonkey. The results are summarized in table 2. As expected, highly significant positive selection (dN/dS = 4.0 in 16.3% of the amino acid residues, *P* value < 0.0001) was detected for the entire primate clade. In contrast, no evidence of positive selection was detected for individual dermopteran species nor for the internal branches of the dermopteran clade (fig. 4; nodes I and II). Importantly, Datamonkey detected strong positive selection (dN/dS = 12.6 in 10.4% of the amino acid residues,

P value = 0.0007) in the MRCA of the dermopteran clade (fig. 4; node III). PAML analysis yielded similar results with similar significance levels.

Next, we tried to locate the individual sites under positive selection in the dermopteran TRIM5α. We calculated the dN/dS values for each position across the entire TRIM5α alignment (supplementary fig. S4, Supplementary Material online). Clusters of residues with high dN/dS values clearly overlap with the V1–V4 variable regions in the B30.2 domain. These variable regions are known to determine the virus specificity of the TRIM5α restriction activity (Yap et al. 2005; James et al. 2007). Further, individual TRIM5α residues subjected to statistically significant positive selection were detected using the REL (Pond and Frost 2005b) and Mixed Effects Model Evolution (MEME) analyses (Murrell et al. 2012) implemented in Datamonkey (supplementary fig. S4, Supplementary Material online). These analyses identified 16 positively selected TRIM5α residues with P value <0.05 (for MEME) or Bayesian factor >50 (REL). More than 80% of these sites were also confirmed (with posterior probability $>90\%$) by BEB analysis in PAML (Yang et al. 2005). Six out of eight of the positively selected residues detected in the B30.2 domain reside in the variable regions, as depicted in the protein alignment (supplementary fig. S4, Supplementary Material online). Interestingly, at five positively selected positions in the B30.2 domain, the specific amino acid residue was conserved in all dermopteran sequences. This suggests that the positive selection did not act on these residues after the divergence of the extant dermopteran species, but may have acted on their ancestor.

In summary, we obtained strong evidence that *TRIM5* was under positive selection pressure exclusively in the common ancestor of the extant *Dermoptera* and not in the individual dermopteran species. This implies that the selection of the dermopteran *TRIM5* is episodic in nature and could have occurred anytime since the origin of the dermopteran lineage up until the diversification of its extant species. We were not able to specify the timing of this episode more precisely, because no other extant dermopteran species are known. Interestingly, this interval includes the period of ELVgv endogenization. Therefore, ELVgv could have contributed to the observed selection pressure on *TRIM5*. As TRIM5α restricts a broad range of retroviruses (Wolf and Goff 2008), any such retrovirus species that infected the dermopteran ancestor could have also participated in the selection. Indeed, the *Galeopterus* genome assembly contains multiple ERV lineages in addition to ELVgv (data not shown). Further experimental work, including functional assays, would be needed to support the existence of an arms race between the dermopteran *TRIM5* and ELVgv. Such a scenario has been proposed for the RELIK endogenous lentivirus, where virus presence correlates with the positive selection of lagomorph *TRIM5* (Fletcher et al. 2010; de Matos et al. 2011; Yap and Stoye 2013). A previous study of various lemur species showed that the variation of TRIM5α restriction specificity depends on the presence of insertion in the V2 region of the B30.2 domain (Rahm et al. 2011). Interestingly, the V1, V2 and V3 regions of the dermopteran TRIM5α contain several short indels, when compared with primate sequences (supplementary fig. S4, Supplementary Material online). Further analysis of the impact of these mutations on TRIM5α antiretroviral activity could provide valuable insight into the evolution of this restriction factor in the dermopteran lineage.

Conclusion

The analysis of orthologous proviral sequences from several dermopteran species allowed us to study the evolutionary history of the ELVgv endogenous lentivirus residing in the colugo genome. Our findings strongly suggest that this virus represents the oldest evidence of lentiviral existence to date and move the estimated origin of the lentivirus genus substantially deeper into the past. ELVgv was probably active during a broad time interval prior to the divergence of the extant dermopteran species, ~20–60 Ma. We also provide evidence for episodic positive selection acting on the *TRIM5* antiviral restriction factor in the common ancestor of the extant *Dermoptera*. Furthermore, we describe several positively selected residues of the TRIM5α protein which may be involved in its antiviral activity.

Materials and Methods

Whole-Genome Amplification and Validation of Dermopteran Samples

Genomic DNA samples of two *G. variegatus* subspecies (*G. v. peninsulae*, designated as GVA3 and *G. v. variegatus*, designated as GVA5) and of *C. volans* (CVO) were provided by W. Murphy (Texas A&M University). The sample identity was confirmed by PCR amplification of *FES* and *CHRNA1* loci, which were described for these specimens before (Janecka et al. 2008), using Colugo1F/1R and Colugo2F/2R primer pairs, respectively (supplementary table S1, Supplementary Material online). The *CHRNA1* locus was verified by sequencing of the PCR products in all specimens and comparison with previously described sequences. Due to the very low amounts of DNA, the whole genome amplification of samples was performed using the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare), as described by the manufacturer.

PCR Amplification and Sequencing of ELVgv Proviruses and *TRIM5* Orthologs

The ELVgv RT region was amplified using two primer pairs, ELVgvF1/R1 and ELVgvF2/R2 (supplementary table S1, Supplementary Material online), yielding PCR products of ~245 and 215 bp, respectively. The short junctions between ELVgv 3′ end and the dermopteran genomic DNA were amplified with one primer anchored in ELVgv 3′-LTR (primer ColugoLTR) and the second primer anchored in the 3′ flanking region of proviruses ***A***, ***B*** and ***C*** (primers ColugoA1/B1/C1). The PCR products from proviruses ***A***, ***B*** and ***C*** were 359, 337 and 268 bp long, respectively, and their identity was verified by sequencing. The empty pre-integration sites were detected using primer pairs ColugoA1/A2, ColugoB1/B2 and ColugoC1/C2. To amplify the 3-kb regions from 5′ parts the ELVgv proviruses, a semi-nested PCR approach was used:

first amplification was performed with internal viral primer ELVgvR1 and a second primer anchored in the 5′ flanking region of proviruses ***A***, ***B*** and ***C*** (primers ColugoA2/B2/C2). Second PCR was then performed with viral primer ELVgvR2 and the same flanking primer as in the first PCR. In cases when this approach was not successful, variant viral primers ELVseq8 or ELVseq10, closer to the 5′ flank, were used. The PCR products obtained were sequenced; in cases when heterozygosity was detected, the products were subcloned and multiple clones were resequenced.

TRIM5α sequence of *G. variegatus* was reconstructed using the colugo genome draft assembly and high-throughput genomic and RNA-seq data (accession numbers: SRR585693, SRR593622, SRR1261671, and SRR1261617) available at the National Center for Biotechnology Information (NCBI) sequence read archive (SRA). Based on this sequence, the following PCR primers were designed to amplify and sequence *TRIM5α* gene fragments from dermopteran genomic DNA: Trim_f1/Trim_r1 for amplification of exons 2 and 3; Trim_f2/Trim_r2 for exon 4 amplification; Trim_f3/Trim_r3 for amplification of exons 6, 7, and 8.

Datasets

A total of 11 proviral nucleotide sequences were used in evolutionary analysis of ELVgv. These sequences correspond to three sets of orthologous viral integrations in four dermopteran specimens.

To detect the signature of positive selection acting on the dermopteran and primate *TRIM5* orthologs, protein-coding nucleotide sequences with following GenBank accession numbers were used: *Homo sapiens* (NM_033034), *Pan troglodytes* (NM_001012650), *Gorilla gorilla* (AY843510), *Pongo pygmaeus* (AY843513), *Callithrix pygmaea* (AY843512), *Nomascus leucogenys* (DQ229283), *Macaca mulatta* (AY523632), and four dermopteran sequences.

In the phylogenetic analysis of dermopteran and primate TRIM paralogs, following protein-coding nucleotide sequences were used in addition to the *TRIM5* dataset mentioned earlier: *M. mulatta TRIM22* (NM_001113359), *Pongo abelii TRIM22* (NM_001159808), *H. sapiens TRIM22* (NM_006074), *P. troglodytes TRIM22* (NM_001113396), *G. gorilla TRIM22* (XM_004050562), *P. abelii TRIM5* (NM_001131070), *Callithrix jacchus TRIM5* (XM_009007572), *C. jacchus TRIM22* (XM_009007574), *N. leucogenys TRIM22* (NM_001280099), *H. sapiens TRIM34* (NM_001003827), *P. troglodytes TRIM34* (NM_001205177), *C. jacchus TRIM34* (XM_009007571), *N. leucogenys TRIM34* (XM_003254844).

Copy-Number Analysis of Putative Genomic Duplications

For accurate absolute quantification of putatively duplicated genomic loci, ddPCR system QX200 (Bio-Rad, Hercules, CA) was used. Each reaction mixture had a total volume of 20 µl, containing 1× QX200 ddPCR Evagreen Supermix (Bio-Rad), 1 ng of genomic DNA (without whole genome amplification), and 250 nM (each) the forward and reverse primers. The reactions were treated for droplet generation according to the manufacturer's manual and then amplified with the following conditions: 1 cycle of 5 min at 95 °C and then 40 cycles consisting of 15 s at 95 °C and 40 s at 59 °C followed by 1 cycle of 5 min at 72 °C, 5 min at 4 °C and 5 min at 90 °C. Samples were analyzed by droplet reader and QuantaSoft software (Bio-Rad) with thresholds set manually. Primers used for ddPCR (supplementary table S1, Supplementary Material online) were: RbXf and RbXr (control genomic locus in rabbit X chromosome; oryCun2 rabbit genomic assembly, chrX:15,687,175-259); RbYf and RbYr (control genomic locus in rabbit Y chromosome; GenBank: AY785433); R1f and R1r (R1 putative duplication in rabbit genome; GenBank: AAGW02078017); R2f and R2r (R2 putative duplication in rabbit genome; GenBank: AAGW02075161); Chrna1f and Chrna1r (*CHRNA1* control locus in dermopteran genome). Two putative duplications of ELVgv integration sites (Hron et al. 2014) were analyzed; region G1 (GenBank: JMZW01362493) and region G2 (GenBank: JMZW01477984). For each region, two ddPCR primer pairs were used: one primer pair overlapped the LTR-host DNA junctions (G1f1 and G1r1 for region G1, G2f1 and G2r1 for region G2), the second primer pair was located outside the virus sequences in the flanking regions (G1f2 and G1r2 for region G1, G2f1 and G2r2 for region G2).

Phylogenetic Analysis

Nucleotide sequences were aligned using the MUSCLE algorithm implemented in MEGA 6 software (Tamura et al. 2013) and checked manually. ML phylogeny was generated using MEGA 6. All positions containing gaps and missing data were excluded. This resulted in dataset with a total of 571 positions for alignment of 11 ELVgv proviral sequences, 1278 positions for alignment of 24 *TRIM5/22/34* sequences, and 1305 positions for alignment of 11 *TRIM5α* sequences. Kimura 2-parameter (K2P) model with uniform rates among sites was used for all datasets. Bootstrap support for each node was evaluated with 1000 replicates.

Bayesian-based phylogenetic tree was reconstructed using BEAST2 software (Bouckaert et al. 2014). K2P model with gamma distribution (4 categories) of rates among sites was used as a substitution model, and Yule speciation model was used as a tree prior. Markov chain Monte Carlo (MCMC) run with 10 millions of generations was done. Trees were sampled every 1000 generations. Effective Sample Size (ESS) values of all parameters were >200. Consensus tree was constructed using TreeAnnotator software included in BEAST2 package.

Substitution Rate Estimation and LTR Aging

Substitution rates for specific loci in the dermopteran genome were estimated using calibration constraints assuming the *Galeopterus* and *Cynocephalus* split time as 15.3 Ma (Hedges et al. 2006). Specifically, the genetic distance between orthologous sequences in *Cynocephalus* and *Galeopterus* species were calculated using the K2P model with uniform rates among sites. Substitution rates were determined by dividing the average *Galeopterus*/*Cynocephalus* genetic distance for each locus by 2 times multiplied *Galeopterus* and *Cynocephalus* split time. These substitution rates were used

for the proviral integration dating by the LTR-aging method (Johnson and Coffin 1999).

Time-Calibrated Phylogeny of ELVgv Proviruses

Molecular dating analysis of proviral sequences was performed either by using the ML-based RelTime method implemented in MEGA 6 (Tamura et al. 2013) or by Bayesian method with BEAST2 software (Bouckaert et al. 2014). Relative times were optimized and converted to absolute divergence times based on the calibration constraints assuming *Galeopterus*/*Cynocephalus* split time to 15.3 Ma (Hedges et al. 2006). In the ML approach, the same parameters as in phylogeny reconstruction were used. The suitability of molecular clock assumption was not rejected at a 5% significance level by comparing the ML value for the given topology with and without the molecular clock constraints.

In the BEAST analysis, the sequence evolution was assumed to follow relaxed molecular clock with lognormal distribution, which was shown as an appropriate model for this type of analysis (Drummond et al. 2006). K2P with gamma distribution (4 categories) of rates among sites were used as a substitution model. Yule speciation process was used as a tree prior. Markov chain Monte Carlo (MCMC) run with 10 millions of generations was performed with sampling every 1000 generations. ESS values of all parameters were >200.

Lineage-Specific Analysis of Positive Selection

Overall dN/dS for each branch, so-called branch model (Yang 1998; Yang and Nielsen 1998), of TRIM5α phylogeny was determined using codeml program of the PAML 4.7 package (Yang 2007) and REL method in the Datamonkey web server (Pond and Frost 2005a; Delport et al. 2010). For codeml program, several simulations of a free-ratio model were performed with F3X4 codon frequency model and multiple seed dN/dS values to ensure convergence of parameter optimization. In datamonkey analysis, Codon Model Selection (CMS) module was used to select the best fitting substitution model. The dN/dS variation across branches was also confirmed by comparing the one-ratio model with the free-ratio model in PAML (data not shown).

In the further analysis, the model assuming several groups of residues specified by different dN/dS for each branch was employed (branch-site model). Presence of positive selection acting on particular branches was tested with branch-site model of positive selection in codeml program (Zhang et al. 2005). Specifically, modified branch-site model A (model = 2 NSsites = 2) was compared with the null model (dN/dS value is fixed to 1) using Likelihood Ratio Test (LRT). *P* values corresponding to positive selection significance were calculated based on LRT statistics. Alternatively, branch-site REL test for episodic positive selection (Pond et al. 2011) in Datamonkey web server was employed. Similarly to PAML analysis, LRT statistics was used to evaluate significance of positive selection acting on particular branches of TRIM5alpha phylogeny. The *P* values were corrected for multiple testing using the Holm–Bonferroni method.

Site-Specific Analysis of Positive Selection

The REL method (Pond and Frost 2005b) and MEME method (Murrell et al. 2012) implemented in the Datamonkey web server were used to detect positively selected residues in the TRIM5α protein sequence. REL is an extension of a well-known codon-based selection analysis. MEME is a more specialized codon-based method to detect episodic selection pressure (Murrell et al. 2012). *P* values representing significance of positive selection for each site were calculated using LRT statistics. BEB positive selection analysis (Yang et al. 2005) implemented in PAML 4.7 was performed under M2a site model and used as a validation of the REL method.

Supplementary Material

Supplementary figures S1–S4 and table S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank William Murphy for providing the dermopteran genomic DNA samples. This work was supported by the Czech Ministry of Education, Youth and Sports under the program NÁVRAT (LK11215). The work was also institutionally supported by RVO:68378050.

References

Aswad A, Katzourakis A. 2012. Paleovirology and virally derived immunity. *Trends Ecol Evol*. 27:627–636.

Bouckaert R, Heled J, Kuehnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comput Biol*. 10(4):e1003537.

Cui J, Holmes EC. 2012. Endogenous lentiviruses in the ferret Genome. *J Virol*. 86:3383–3385.

de Matos AL, van der Loo W, Areal H, Lanning DK, Esteves PJ. 2011. Study of Sylvilagus rabbit TRIM5 alpha species-specific domain: how ancient endoviruses could have shaped the antiviral repertoire in Lagomorpha. *BMC Evol Biol*. 11:294.

Delport W, Poon AFY, Frost SDW, Pond SLK. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26:2455–2457.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 4:699–710.

Fletcher AJ, Hue S, Schaller T, Pillay D, Towers GJ. 2010. Hare TRIM5 alpha restricts divergent retroviruses and exhibits significant sequence variation from closely related lagomorpha TRIM5 genes. *J Virol*. 84:12463–12468.

Gifford RJ, Katzourakis A, Tristem M, Pybus OG, Winters M, Shafer RW. 2008. A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proc Natl Acad Sci U S A*. 105:20362–20367.

Gilbert C, Maxfield DG, Goodman SM, Feschotte C. 2009. Parallel germline infiltration of a lentivirus in two malagasy lemurs. *PLoS Genet*. 5(3):e1000425.

Han G-Z, Worobey M. 2012. Endogenous lentiviral elements in the Weasel family (Mustelidae). *Mol Biol Evol*. 29:2905–2908.

Han G-Z, Worobey M. 2015. A primitive endogenous lentivirus in a colugo: insights into the early evolution of lentiviruses. *Mol Biol Evol*. 32:211–215.

Hayward A, Cornwallis CK, Jern P. 2015. Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. *Proc Natl Acad Sci U S A*. 112:464–469.

Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22:2971–2972.
Hodgkinson A, Eyre-Walker A. 2011. Variation in the mutation rate across mammalian genomes. *Nat Rev Genet*. 12:756–766.
Hron T, Fabryova H, Paces J, Elleder D. 2014. Endogenous lentivirus in Malayan colugo (*Galeopterus variegatus*), a close relative of primates. *Retrovirology* 11:6.
James LC, Keeble AH, Khan Z, Rhodes DA, Trowsdale J. 2007. Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. *Proc Natl Acad Sci U S A*. 104:6200–6205.
Janecka JE, Helgen KM, Lim NTL, Baba M, Izawa M, Murphy WJ. 2008. Evidence for multiple species of *Sunda colugo*. *Curr Biol*. 18:R1001–R1002.
Janecka JE, Miller W, Pringle TH, Wiens F, Zitzmann A, Helgen KM, Springer MS, Murphy WJ. 2007. Molecular and genomic data identify the closest living relative of primates. *Science* 318:792–794.
Jern P, Coffin JM. 2008. Effects of retroviruses on host genome function. *Annu Rev Genet*. 42:709–732.
Jha AR, Nixon DF, Rosenberg MG, Martin JN, Deeks SG, Hudson RR, Garrison KE, Pillai SK. 2011. Human endogenous retrovirus K106 (HERV-K106) was infectious after the emergence of anatomically modern humans. *PLoS One* 6(5):e20234.
Jha AR, Pillai SK, York VA, Sharp ER, Storm EC, Wachter DJ, Martin JN, Deeks SG, Rosenberg MG, Nixon DF, et al. 2009. Cross-sectional dating of novel haplotypes of HERV-K 113 and HERV-K 115 indicate these proviruses originated in Africa before *Homo sapiens*. *Mol Biol Evol*. 26:2617–2626.
Johnson WE. 2015. Endogenous retroviruses in the genomics era. *Annu Rev Virol*. 2:135–159.
Johnson WE, Coffin JM. 1999. Constructing primate phylogenies from ancient retrovirus sequences. *Proc Natl Acad Sci U S A*. 96:10254–10260.
Kamath PL, Elleder D, Bao L, Cross PC, Powell JH, Poss M. 2014. The population history of endogenous retroviruses in mule deer (*Odocoileus hemionus*). *J Hered*. 105:173–187.
Katzourakis A, Gifford RJ, Tristem M, Gilbert MT, Pybus OG. 2009. Macroevolution of complex retroviruses. *Science* 325:1512.
Katzourakis A, Tristem M, Pybus OG, Gifford RJ. 2007. Discovery and analysis of the first endogenous lentivirus. *Proc Natl Acad Sci U S A*. 104:6261–6265.
Keckesova Z, Ylinen LMJ, Towers GJ, Gifford RJ, Katzourakis A. 2009. Identification of a RELIK orthologue in the European hare (*Lepus europaeus*) reveals a minimum age of 12 million years for the lagomorph lentiviruses. *Virology* 384:7–11.
Kijima TE, Innan H. 2010. On the estimation of the insertion time of LTR retrotransposable elements. *Mol Biol Evol*. 27:896–904.
Li WH, Tanimura M, Sharp PM. 1988. Rates and dates of divergence between aids virus nucleotide-sequences. *Mol Biol Evol*. 5:313–330.
Martins H, Villesen P. 2011. Improved integration time estimation of endogenous retroviruses with phylogenetic data. *PLoS One* 6(3):e14745.
McCarthy KR, Kirmaier A, Autissier P, Johnson WE. 2015. Evolutionary and functional analysis of old world primate TRIM5 reveals the ancient emergence of primate lentiviruses and convergent evolution targeting a conserved capsid interface. *PLoS Pathog*. 11(8):e1005085.
Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet*. 8(7):e1002764.
Perelman P, Johnson WE, Roos C, Seuanez HN, Horvath JE, Moreira MAM, Kessing B, Pontius J, Roelke M, Rumpler Y, et al. 2011. A molecular phylogeny of living primates. *PLoS Genet*. 7(3):e1001342.
Perez-Caballero D, Hatziioannou T, Yang A, Cowan S, Bieniasz PD. 2005. Human tripartite motif 5 alpha domains responsible for retrovirus restriction activity and specificity. *J Virol*. 79:8969–8978.
Pond SLK, Frost SDW. 2005a. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21:2531–2533.
Pond SLK, Frost SDW. 2005b. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol*. 22:1208–1222.
Pond SLK, Murrell B, Fourment M, Frost SD, Delport W, Scheffler K. 2011. A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol*. 28:3033–3043.
Rahm N, Yap M, Snoeck J, Zoete V, Munoz M, Radespiel U, Zimmermann E, Michielin O, Stoye JP, Ciuffi A, et al. 2011. Unique spectrum of activity of prosimian TRIM5 alpha against exogenous and endogenous retroviruses. *J Virol*. 85:4173–4183.
Sawyer SL, Emerman M, Malik HS. 2007. Discordant evolution of the adjacent antiretroviral genes TRIM22 and TRIM5 in mammals. *PLoS Pathog*. 3:e197.
Sawyer SL, Wu LI, Emerman M, Malik HS. 2005. Positive selection of primate TRIM5 alpha identifies a critical species-specific retroviral restriction domain. *Proc Natl Acad Sci U S A*. 102:2832–2837.
Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. 2004. The cytoplasmic body component TRIM5 alpha restricts HIV-1 infection in Old World monkeys. *Nature* 427:848–853.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 30:2725–2729.
Tonjes RR, Niebert M. 2003. Relative age of proviral porcine endogenous retrovirus sequences in *Sus scrofa* based on the molecular clock hypothesis. *J Virol*. 77:12363–12368.
van der Loo W, Abrantes J, Esteves PJ. 2009. Sharing of endogenous lentiviral gene fragments among leporid lineages separated for more than 12 million years. *J Virol*. 83:2386–2388.
Wolf D, Goff SP. 2008. Host restriction factors blocking retroviral replication. *Annu Rev Genet*. 42:143–163.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.
Yang ZH. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*. 15:568–573.
Yang ZH, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol*. 46:409–418.
Yang ZH, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol*. 19:908–917.
Yang ZH, Nielsen R, Goldman N, Pedersen AMK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449.
Yang ZH, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol*. 22:1107–1118.
Yap MW, Nisole S, Stoye JP. 2005. A single amino acid change in the SPRY domain of human Trim5alpha leads to HIV-1 restriction. *Curr Biol*. 15:73–78.
Yap MW, Stoye JP. 2013. Apparent effect of rabbit endogenous lentivirus type K acquisition on retrovirus restriction by lagomorph Trim5 α s. *Philos Tran R Soc B Biol Sci*. 368:20120498.
Zhang JZ, Nielsen R, Yang ZH. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. 22:2472–2479.

Discovery of the first endogenous Deltaretrovirus, in the genome of long-fingered bats (Chiroptera: Miniopteridae)

Helena Farkašová[a,1], Tomáš Hron[a,1], Jan Pačes[a], Pavel Hulva[b,c], Petr Benda[b,d], Robert Gifford[e,2], Daniel Elleder[a,2]

[a]Institute of Molecular Genetics, The Czech Academy of Sciences, Videnska 1083, Prague, 14220, Czech Republic [b]Department of Zoology, Charles University, Vinicna 7, Prague, 12844, Czech Republic [c]Department of Biology and Ecology, University of Ostrava, Chitussiho 10, Ostrava, 71000, Czech Republic [d]Department of Zoology, National Museum (Natural History), Vaclavske nam. 68, Prague, 11579, Czech Republic [e]MRC-University of Glasgow, Centre for Virus Research, 464 Bearsden Road, Glasgow, G12 8TA, United Kingdom [1]Authors contributed equally [2]Corresponding authors

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Retroviruses can create endogenous forms upon infiltration into the germline cells of their hosts. These forms are then vertically transmitted and can be considered as genetic fossils of ancient viruses. All retrovirus genera, with the exception of deltaretroviruses, have had their representation identified in the host genome as a virus fossil record. Here we describe the first endogenous Deltaretrovirus, identified in the germline of long-fingered bats (Miniopteridae). A single, heavily deleted copy of this retrovirus has been found in the genome of Miniopterid species, but not in the genomes of the phylogenetically closest bat families, Vespertilionidae and Cistugonidae. Therefore, the endogenization occurred in a time interval between 20 and 45 million years ago. This discovery closes the last major gap in the retroviral fossil record and provides important insights into the history of deltaretroviruses in mammals.

Deltaretroviruses | Endogenous retroviruses | Chiroptera

Introduction

Deltaretroviruses are a highly unusual genus of retroviruses (family Retroviridae) that have only been identified in a restricted subset of mammalian species. They include the primate T-cell lymphotropic viruses (PTLVs) that infect apes (including humans) and Old World monkeys, and bovine leukemia virus (BLV) that infects cattle. Deltaretrovirus infections are usually asymptomatic, but can cause inflammatory and malignant disease over the longer term. For example, in humans, infection with human T-lymphotropic virus type 1 (HTLV-1) can cause adult T-cell lymphoma/leukemia or HTLV-1 associated myelopathy/tropical spastic paraparesis (1, 2). In cattle, BLV can cause persistent lymphocytosis or leukemia/lymphoma (3, 4).

Understanding of deltaretroviruses is limited by the lack of an endogenous 'fossil record' for this genus (5, 6). Retroviruses are distinguished by a replication strategy in which a DNA copy of the viral genome (a form called a 'provirus') is integrated into the nuclear genome of the host cell. Consequently, retroviral infection of germline cells can lead to retroviral proviruses being vertically inherited as host alleles, called endogenous retroviruses (ERVs). Vertebrate genomes typically contain thousands of ERVs, many of which are derived from retroviruses that circulated millions of years ago. These sequences constitute a partial historical record of the retroviruses that have been encountered by vertebrate species during their evolution (7). However, despite many vertebrate genomes having been sequenced, ERVs derived from deltaretroviruses have yet to be identified. Here we describe the first endogenous Deltaretrovirus, identified in the genome of long-fingered bats (Miniopteridae).

Results

While systematically screening mammalian genomes for ERVs (8) we detected a sequence in the genome of the Natal long-fingered bat (*Miniopterus natalensis*)(9) that disclosed highly significant similarity to Deltaretrovirus Gag proteins. This sequence was identified in a single large contig (GenBank accession no. LDJU01000221, 2.6 megabase-long), and was flanked by the paired long terminal repeat (LTR) sequences characteristic of retroviral proviruses. A 6-bp target site duplication (TSD) sequence (GCCCCC) was identified immediately upstream and downstream of the proviral insertion. We performed manual analysis of raw reads from published *M. natalensis* sequencing projects to accurately recover this proviral locus (see Methods). In addition, we used PCR to confirm the presence of the provirus in the *M. natalensis* genome, as well as in four other *Miniopterus* species. Complete proviruses from all five *Miniopterus* species were sequenced and submitted to GenBank (accession numbers KY250075-9). We named the provirus *Miniopterus* endogenous retrovirus (MINERVa).

The orthologous proviruses obtained from the five miniopterid species were almost identical, differing only by several substitutions and indels. For description of the provirus, we generated a majority rule consensus sequence. The consensus MINERVa genome comprises a 1,789-base-pair (bp) internal region flanked by 604-bp LTRs (Fig. 1). Notably, the entire MINERVa sequence exhibits the characteristic nucleotide composition bias of deltaretroviruses (10), with cytosine (C) strongly overrepresented (Fig. 1). A primer binding site (PBS) specific for proline tRNA is present in the internal region, immediately downstream

Significance

Retroviruses copy their RNA genome into complementary DNA, which is then inserted into the host chromosomal DNA as an obligatory part of their life cycle. Such integrated viral sequences, called proviruses, are passed to the infected cell progeny upon cellular division. If germline cells are targeted, the proviruses become vertically inherited as other host genes, and are called endogenous retroviruses. Deltaretroviruses, which include important human and veterinary pathogens (HTLV-1 and BLV), are the last retroviral genus for which endogenous forms were not known. We have identified the first case of endogenous Deltaretrovirus, which entered the genome of long-fingered bat ancestors more than 20 million years ago. This finding opens the way for elucidating the deep evolutionary history of deltaretroviruses.

Reserved for Publication Footnotes

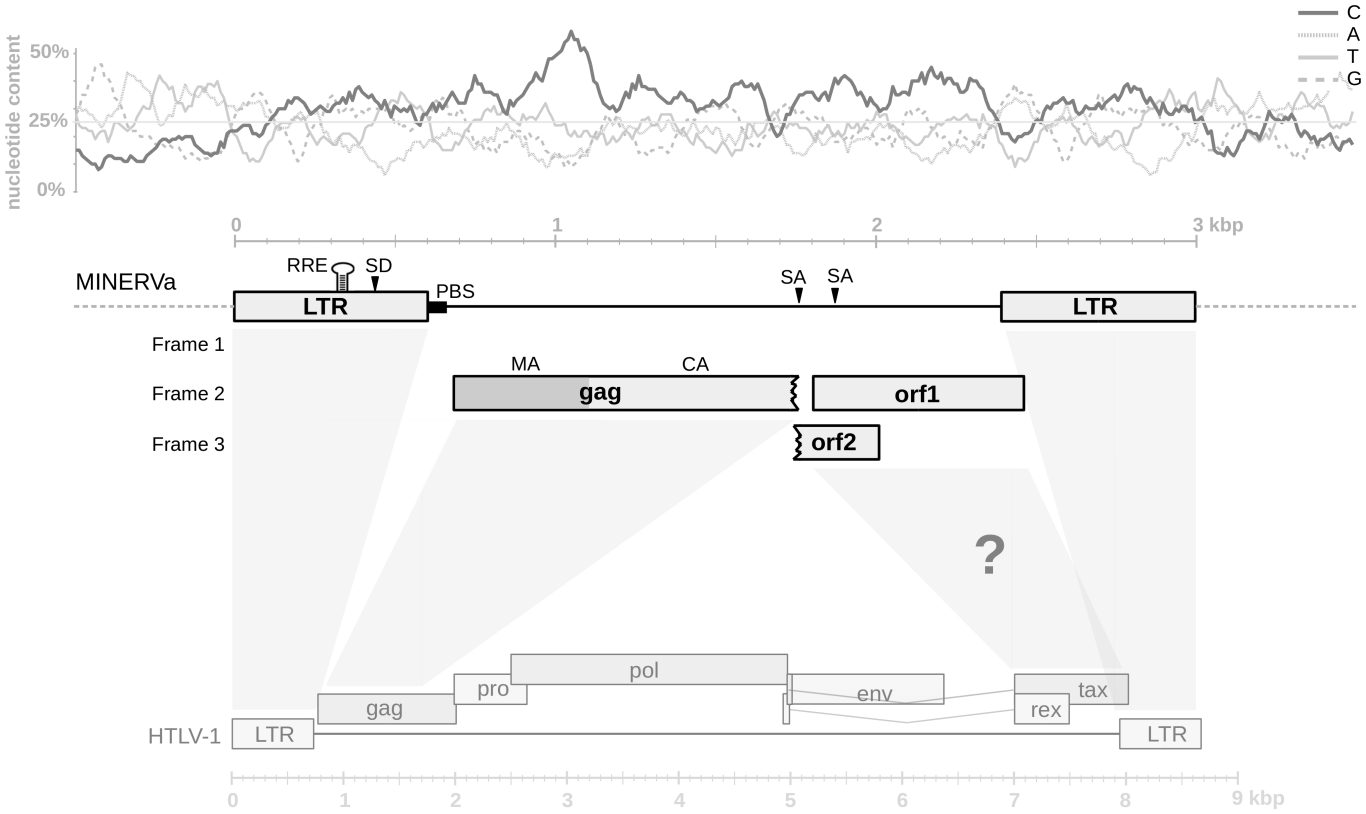

Fig. 1. Genome organization of MINERVa.The consensus sequence of MINERVa is shown schematically in scale. The position of open reading frames and other genomic features are indicated. The comparison with HTLV-1 genome structure is shown (for clarity, some HTLV-1 genes were omitted). The question mark indicates the putative accessory gene region. Nucleotide composition of MINERVa sequence and of 500-bp flanking regions is plotted in scale above the proviral scheme (calculated in 100-bp long windows with 10% overlaps). LTR, long terminal repeat; MA, matrix; CA, capsid; RRE, rex response element; PBS, primer binding site; SD, splice donor site; SA, splice acceptor site.

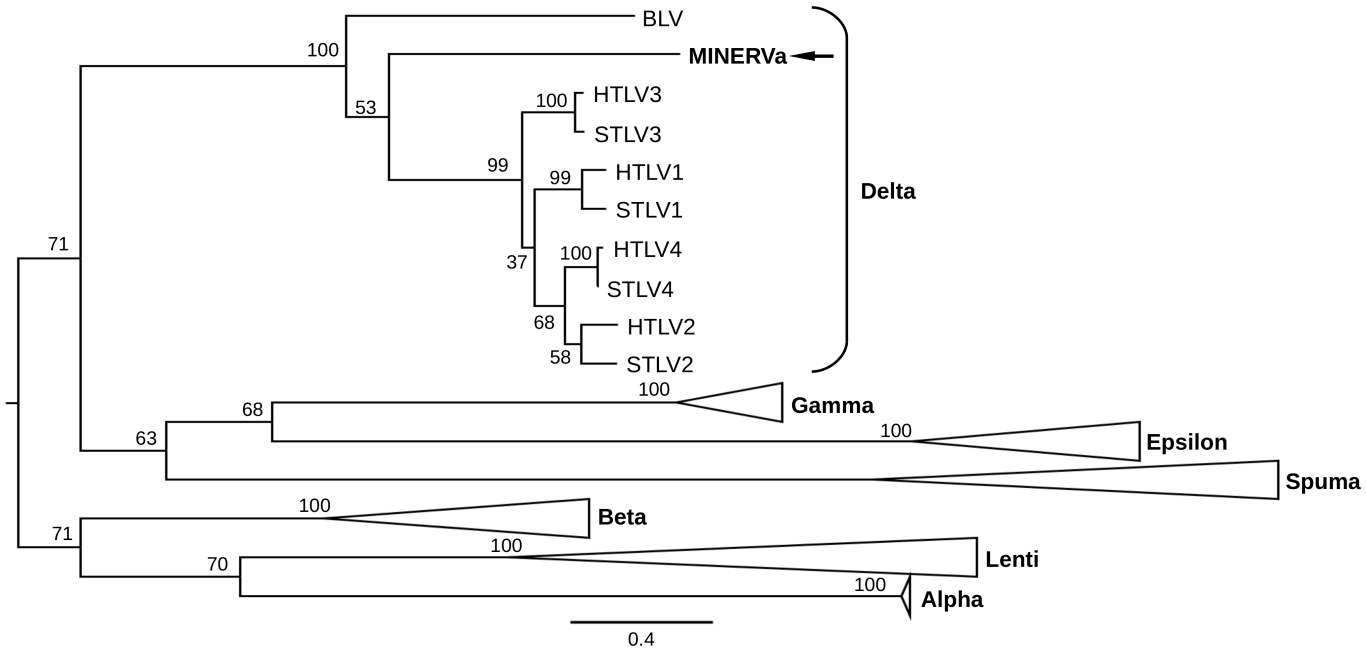

Fig. 2. Phylogenetic relationship of MINERVa to other retroviruses.The ML phylogeny of partial Gag amino acid sequences is shown. Bootstrap supports for each branch are depicted. An arrow highlights position of MINERVa consensus sequence. Branches corresponding to the viruses of particular retroviral genera are collapsed, with the exception of deltaretroviruses. Scale bar indicates number of amino acid substitutions per site.

of the 5' LTR. The internal region encodes three relatively long open reading frames (ORFs), the first of which is 376 amino acids (aa) in length, and encodes a truncated *gag* gene containing putative matrix (MA) and capsid (CA) proteins (Fig. 1 and Fig. S1). The predicted MA and CA amino acid sequences showed 30.8% and 45.4% identity (44.5 and 60.2% similarity) respectively

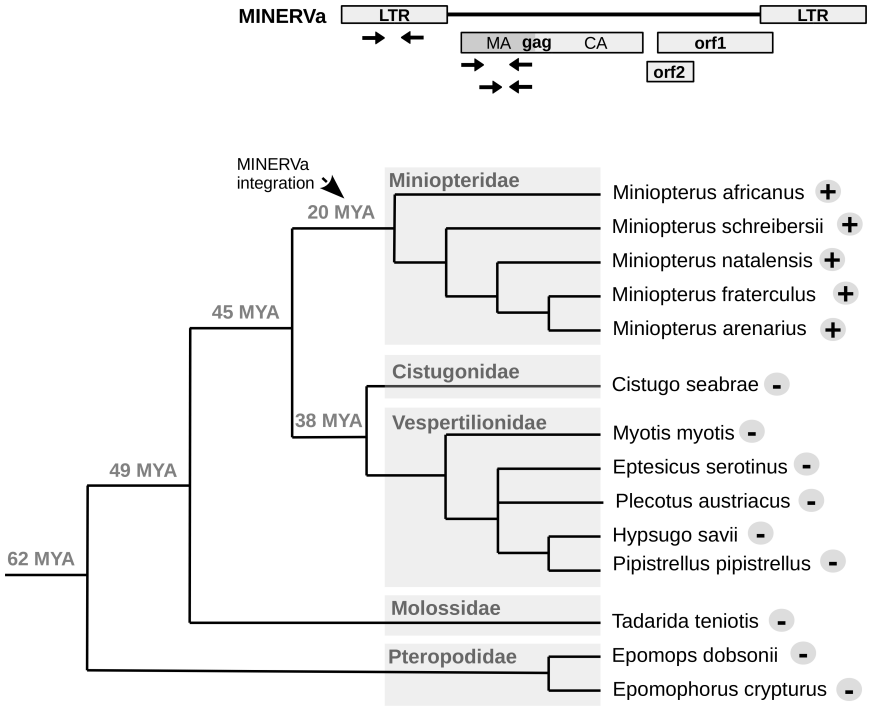

Fig. 3. The presence of MINERVa in various species of bats.Three primer pairs used for PCR screening of selected bat DNA samples are depicted in MINERVa schematics. Results of the screen are shown in a phylogenetic tree of bat species. Plus sign next to the species name indicates positivity for all three MINERVa amplicons; minus sign indicates negativity for all three amplicons. Dating of split of selected species are shown next to the branch nodes (timetree.org estimates (15)).

to those of HTLV-1. The entire putative *gag* ORF was translated and the resulting sequence aligned with Gag proteins of representative retroviruses. In phylogeny constructed using this alignment, the MINERVa Gag grouped robustly within the Deltaretrovirus clade, forming a basal branch separate from both BLV and PTLV group (Fig. 2).

The second and third long ORFs (ORF1 and ORF2 in Fig. 1) are 220 and 88 aa long, respectively, both beginning immediately downstream of the truncated *gag* ORF. They overlap each other and ORF1 is extending a short distance (76 bp) into the 3' LTR. We could not detect homology to any of the genetic elements that would be expected to occur immediately downstream of capsid - nucleocapsid (NC), protease (PR), polymerase (*pol*) and envelope (*env*) - or indeed to any known proteins. One way this result might be expected would be if a region spanning from end of *gag*, through PR and *pol* and up to the end of *env* had been deleted (see Fig. 1). A deletion event of this nature would leave behind a truncated *gag* gene, plus the 3' region of the provirus genome, which is situated downstream of *env* and contains multiple overlapping accessory genes. The absence in MINERVa of any homology to the relatively conserved transmembrane (TM) domain, which occurs toward the C-terminus of the Env protein, is consistent with this having occurred. In all previously characterized deltaretroviruses, the 3' region of the genome encodes accessory genes, including the regulators of gene expression *tax* and *rex*, in addition to other, less well characterized genes (Fig. 1)(11).

Additional, independent lines of evidence support the inference that ORF1/2 are accessory genes derived from the 3' region of the ancestral MINERVa genome. Firstly, we detected the characteristic nucleotide bias of deltaretroviruses across the entire MINERVa genome, including the region that encodes ORF1/2, consistent with this region being derived from an ancestral Deltaretrovirus (as opposed to insertion of genome-derived sequence into the MINERVa provirus following its integration, for example). The MINERVa genome also exhibited features that were consistent with ORF1/2 being *bona fide* retroviral genes with a role in regulation of gene expression. These included: (i) The presence of a leucine-rich region toward the C-terminus of the predicted *orf1* gene product, as is found in the *tax* gene of primate and bovine deltaretroviruses (12, 13); (ii) The presence of a predicted stem-loop (Fig. S2) in the region of the LTR that would be expected to encode a Rex-responsive element (presuming that MINERVa replication was regulated in a similar post-transcriptional manner to other deltaretroviruses), and (iii) The fact that ORF1/2 overlap, resembling the *tax*/*rex* overlap in other deltaretroviruses (12, 14).

We performed a series of analyses to determine the distribution of the MINERVa sequences in bats. From our *in silico* genome screening it was already clear that the sequence was absent in more distantly related bat groups, such as Old World fruit bats (family Pteropodidae), as well in some more closely related lineages (i.e. *Myotis* spp.). To investigate the presence of MINERVa more thoroughly, we performed a PCR-based screen of the phylogenetically closest bat species. In addition to the *Miniopterus* specimens already mentioned, samples were obtained from nine bat species from other families (Vespertilionidae, Cistugonidae, Molossidae and Pteropodidae) (Table S1). The species designation of these samples was confirmed by amplification and phylogenetic analysis of *cytB* or *rag2* genes (see Methods). PCR primers designed to target three different amplicons in provirus (Table S2) confirmed that the MINERVa insertion is present in all miniopterid bat species examined (Fig. 3). MINERVa could be detected neither in species belonging to the most closely related bat families (Vespertilionidae and Cistugonidae) nor in more distant species.

To confirm that only a single MINERVa integration is present in all five of the miniopterid species examined (i.e. two alleles per diploid genome), we used digital droplet PCR (Fig. S3). All these single-copy MINERVa insertions in distinct *Miniopterus* species are clearly orthologous, sharing >124 bp of homologous flanking sequence on either site of the proviral integration site. The five *Miniopterus* bats included in our study are estimated to have diverged ~20 million years ago (MYA) (15, 16), establishing that genomic infiltration occurred prior to this date. The absence of related virus sequences in all members of the sister families Ves-

pertilionidae and Cistugonidae examined indicates that invasion is unlikely to have occurred more than 45 MYA (17). Thus, we estimate that MINERVa entered the germline of miniopterid bat ancestors at some point in the period spanning 45-20 MYA (Fig. 3).

An alternative approach for estimating the age of proviral insertions is to determine the divergence between paired LTRs (which are identical at the time of integration) and apply a molecular clock (18). Taking in account the fact that ERV integration precedes the split of the host species, each proviral LTR pair should contain more changes (which accumulate since the time of integration) than orthologous LTR sequences from different species (which accumulate mainly after the species divergence). Analysis of MINERVa sequence from individual miniopterid species, however, showed that the divergence of LTR pairs in each proviral sequence is much lower than divergence between orthologous LTR sequences from some of the species analyzed (e.g. 5'LTR form *M. schreibersii* is more similar to its 3'LTR counterpart than to 5'LTR of *M. fraterculus*) (Fig. S4). This observation provides compelling evidence that multiple gene conversion events - a phenomenon that has previously been described in ERVs (19, 20) - have occurred between the 5' and 3' LTRs of individual proviruses. This fact is precluding the LTR-based approach of age estimation. In addition to phylogenetic evidence for gene conversion, we identified a 5-bp insertion that was unique to *M. schreibersii* MINERVa provirus, but present in both LTRs. This pattern of variation is extremely unlikely to be accounted for by any process other than gene conversion between the 5' and 3' LTRs.

The divergence of internal proviral sequences (excluding both LTRs) cannot yield a time estimate for virus integration. However, like other genomic loci, it should reflect the changes accumulated since the split of the miniopterid species analyzed. *M.africanus* split from the other miniopterid species around 20 MYA (Fig. 3). The average sequence divergence of *M. africanus* MINERVa provirus to its orthologs in other miniopterid species was found to be 1.31 ± 0.35% (mean ± standard deviation), which corresponds to a substitution rate of 0.66 ± 0.18 substitutions/nucleotide/year. This falls within the range of mammalian neutral substitution rate estimates (21, 22).

Given the predicted age of the insertion, it was intriguing that the *gag* ORF was intact in five out of the six MINERVa alleles. However, multiple simulations of MINERVa *gag* neutral evolution recapitulated this situation in 20% of cases (1,000,000 replicates; average number of sequences with at least one stop-codon = 2.52 out of 6). Thus, there is no strong evidence of selection for *gag* coding sequence conservation, although this approach only considers nucleotide substitutions and not indels (23).

The genomic locus in *M. natalensis* where MINERVa integrated does not contain any predicted genes. Orthologous loci, without the provirus, could be detected in several of the published bat genomes. The chromosomal location of the MINERVa locus could not be determined, because none of the bat genomes is yet mapped to chromosomes (24).

Discussion

Exogenous retroviruses have been grouped into seven genera, only five of which are known to infect mammals. The discovery of MINERVa means that endogenous fossils have now been identified in mammalian genomes for all five of these retroviral genera. However, the representation of these five genera in the retroviral fossil record is very uneven. ERVs derived from retroviruses with simpler genome structures (Gammaretrovirus, Betaretrovirus) are relatively common, whereas only a handful have been identified for the Lentivirus, Spumavirus, and (now) Deltaretrovirus genera (5, 6). This could reflect inefficient entry of these viruses into germline cells, or inherent barriers to their germline replication (e.g. toxicity of gene products) (7). Notably, only a single MINERVa copy was identified. Thus, one possibility is that MINERVa was generated with the same, highly deleted genome structure that we see in all present day copies, and being effectively 'dead on arrival', was fixed without any virus-driven increase in germline copy number (as is presumed to occur for endogenous sequences derived from non-retroviral viruses).

Deltaretrovirus is perhaps the most enigmatic of the five retroviral genera that infect mammals, and the discovery of MINERVa is therefore particularly illuminating. Firstly, it provides the first unequivocal evidence that this genus has a truly ancient origin in mammals. We identified orthologous copies in miniopterid species that are estimated to have diverged ~20 MYA, establishing that deltaretroviruses have been infecting mammals for at least this long. Previous studies have demonstrated that Deltaretrovirus infection in humans likely predates the last Ice Age (25), but MINERVa provides the first unequivocal evidence that Deltaretrovirus infection has impacted mammals during a substantial part of their evolution.

The calibration of Deltaretrovirus evolution through the identification of a fossil sequence reveals that the characteristic features of this genus had already evolved by the early Miocene (~23-16 MYA). These include the marked nucleotide-bias that is a hallmark of Deltaretrovirus genomes (10). Nucleotide biases are a feature of many retroviral genomes, but deltaretroviruses stand apart from all other retroviral genera in having C-rich genomes. The biological significance of these biases is uncertain, but the stability of this feature across Deltaretrovirus evolution suggests it represents an adaptation of some kind.

Complex regulation of genome expression is another characteristic feature of deltaretroviruses. Analysis of the MINERVa genome indicated that the ancestral progenitor likely encoded a region with accessory genes. The putative ORFs we identified in this region did not disclose homology to the Tax and Rex proteins of exogenous deltaretroviruses, but since these genes are relatively poorly conserved, this might be expected.

The discovery of MINERVa extends the known host range of the Deltaretrovirus genus to a new mammalian order (Chiroptera). It also raises the questions about the role of bats in Deltaretrovirus evolution. Traits associated with movement capacity are especially pronounced in miniopterids. Their most apparent apomorphy, elongated wings, presumably enabled them to colonize almost all tropic and sub-tropic regions of the Old World and become one of the most widespread mammalian genera (26, 27). They also concentrate in mass roosts of thousands of individuals in caves with high humidity, which could facilitate virus transmission. Conceivably, deltaretroviruses may infect bats in the present day. Searches of available metagenome and transcriptome datasets did not reveal any matches to MINERVa or other deltaretroviruses, but these data represent a relatively limited sample.

In conclusion, the identification MINERVa provides important insights into Deltaretrovirus evolution. It also fills a major gap in ERV record by eliminating the last retrovirus genus for which endogenous forms were not known.

Methods

Next Generation Sequence Data Analysis. Sequence datasets available from the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) from miniopterid species genome or transcriptome (PRJNA270665, PRJNA270639 and PRJNA218524) were queried by BLAST (28) or downloaded and analyzed using CLC genomics workbench 9.5 (http://www.clcbio.com) or DNASTAR Lasergene 10.0.0 (http://dnastar.com). This initial analysis was used to correct errors in the original MINERVa-containing contig from the *M. natalensis* genome assembly.

Samples from Bats. The bat tissue samples were obtained from museum specimens (National Museum Prague) as parts of the pectoral muscles and from released bats caught during various ecological studies as wing punch biopsies stored in genetic bank (Charles University, Prague). The bat species

were identified with respect to their external morphological traits and confirmed by amplification and sequencing of cytochrome b (*cytb*) or recombination activating gene (*rag2*) loci. Total DNA from the ethanol preserved specimens was isolated using phenol-chloroform extraction method.

PCR and Sequencing. The complete MINERVa provirus sequence was PCR-amplified using two strategies (primers listed in **Table S2**): (i) a nested PCR approach with primers anchored in genomic flanking regions (primers F6 and R4 in first round, primers F5 and R5 second round), or (ii) in two overlapping parts using one primer anchored in the genomic flanking region and second primer in the provirus sequence (5' provirus part amplified using primers F6 and R1, 3' provirus part using semi-nested PCR with primers F6 and R4 in first round and F1 and R4 in second round). PCR products were isolated from agarose gels and directly sequenced. The *cytb* locus was amplified using primers cytBMVZ04 and cytBMVZ05 (29), the *rag2* locus using primers 968R and 428F (30). In some *Miniopterus* specimens, the *cytb* locus was amplified using primers CYTB1 and CYTB2 (**Table S2**). To assess the presence of MINERVa sequence in various bat species, two amplicons in the *gag* gene (primers F1 and R1, or F2 and R1) and one amplicon in LTR (F8 and R6) were employed.

Provirus Copy-Number Determination. Digital droplet PCR (ddPCR) system QX200 (Bio-Rad, Hercules, CA) was used to accurately quantify the MINERVa proviral copies in miniopterid samples. Template genomic DNAs were first digested with *Sac*I restriction endonuclease to prevent the occurrence of two LTR sequences in one molecule. The reactions containing 10 ng of DNA were then treated for droplet generation according to the manufacturer's manual and PCR-amplified. The amplified samples were analyzed by droplet reader and QuantaSoft program (Bio-Rad) with thresholds set manually. Primers used for ddPCR (**Table S2**) were: F2 and R1 (MINERVa *gag* region); F8 and R6 (MINERVa LTR); F5 and R6 (5' provirus-genome junction); F4 and R5 (3' provirus-genome junction); F10 and R9 (genomic locus 1); F9 and R8 (genomic locus 2); F7 and R7 (genomic locus 3).

Phylogenetic Analysis. Translated nucleotide sequences of the MINERVa *gag* consensus and other retroviral *gag* sequences were aligned using the MAFFT v7.271 with L-INS-i algorithm (31). Columns containing more than 80% of gaps were discarded resulting in an alignment with a total of 644 positions. Maximum Likelihood (ML) phylogeny was generated using PhyML v3.0 (32). LG model with gamma distribution (four categories) of rates among sites was used as a best fitting substitution model (according to the Akaike Information Criterion calculated in Smart Model Selection module of PhyML). The SPR operations in an optimized BioNJ starting tree were used for searching of the final tree. Bootstrap support for each node was evaluated with 1,000 replicates. The accession numbers of *gag* sequences used are: RSV (NP_056886), ALV (BAK64245), MPMV (NP_056893), JSRV (AAD45224), SIV2 (AAA47561), FLV (NP_955576), MLV (NP_057933), BLV (NP_056897), HTLV-1 (BAA02929), HTLV-2 (AAB59884), HTLV-3 (ACF40912), HTLV-4 (YP_002455784), STLV-1 (AAU34008), STLV-2 (YP_567048), STLV-3 (CAA68892), STLV-4 (AHH34968), VISNA (NP_040839), FIV (NP_040972), HIV (AAB50258), WDSV (AAC82607), WEHV-1 (AAD30047), SFV (NP_056802), BFV (AFR79238). The same software was used for phylogenetic inference of MINERVa LTRs and internal nucleotide sequences. Kimura 2-parameter (K80) model with gamma distribution (four categories) of rates among sites was used as a substitution model. The transition/ transversion ratio was assumed to be 4. SPR operations in optimized BioNJ starting tree were used for searching of the final tree. Bootstrap support for each node was evaluated with 1,000 replicates.

Simulation of Neutral Evolution. The probability of MINERVa *gag* ORF disruption was evaluated by simulating the *gag* sequence evolution using Seq-Gen v1.3.3 (33). The *gag* ancestral sequence and ML phylogeny was inferred from six MINERVa orthologous copies in Miniopteridae. The transition/ transversion ratio was assumed to be 4. The presence of premature stop-codons in simulated *gag* orthologs was counted for 1,000,000 iterations.

Acknowledgements

We are grateful to Vladimír Pečenka for advice on PCR amplifications. This work was supported by the Czech Ministry of Education, Youth and Sports under the program NÁVRAT (LK11215), and also institutionally supported by RVO:68378050. Access to computing and storage facilities was provided by ELIXIR CZ research infrastructure project (MEYS Grant No: LM2015047).

1. Poiesz BJ, *et al.* (1980) Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci U S A* 77(12):7415-7419.
2. Coffin JM (2015) The discovery of HTLV-1, the first pathogenic human retrovirus. *Proc Natl Acad Sci U S A* 112(51):15525-15529.
3. Barez PY, *et al.* (2015) Recent Advances in BLV Research. *Viruses* 7(11):6080-6088.
4. Miller JM, Miller LD, Olson C, & Gillette KG (1969) Virus-like particles in phytohemagglutinin-stimulated lymphocyte cultures with reference to bovine lymphosarcoma. *J Natl Cancer Inst* 43(6):1297-1305.
5. Hayward A, Cornwallis CK, & Jern P (2015) Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. *Proc Natl Acad Sci U S A* 112(2):464-469.
6. Hayward A, Grabherr M, & Jern P (2013) Broad-scale phylogenomics provides insights into retrovirus-host evolution. *Proc Natl Acad Sci U S A* 110(50):20146-20151.
7. Johnson WE (2015) Endogenous Retroviruses in the Genomics Era. *Annu Rev Virol* 2(1):135-159.
8. Hron T, Fabryova H, Paces J, & Elleder D (2014) Endogenous lentivirus in Malayan colugo (Galeopterus variegatus), a close relative of primates. *Retrovirology* 11:84.
9. Eckalbar WL, *et al.* (2016) Transcriptomic and epigenomic characterization of the developing bat wing. *Nat Genet* 48(5):528-536.
10. Kypr J, Mrazek J, & Reich J (1989) Nucleotide composition bias and CpG dinucleotide content in the genomes of HIV and HTLV 1/2. *Biochim Biophys Acta* 1009(3):280-282.
11. Yoshida M (2005) Discovery of HTLV-1, the first human retrovirus, its unique regulatory mechanisms, and insights into pathogenesis. *Oncogene* 24(39):5931-5937.
12. Currer R, *et al.* (2012) HTLV tax: a fascinating multifunctional co-regulator of viral and cellular pathways. *Front Microbiol* 3:406.
13. Aida Y, Murakami H, Takahashi M, & Takeshima SN (2013) Mechanisms of pathogenesis induced by bovine leukemia virus as a model for human T-cell leukemia virus. *Front Microbiol* 4:328.
14. Pavesi A, Magiorkinis G, & Karlin DG (2013) Viral proteins originated de novo by overprinting can be identified by codon usage: application to the "gene nursery" of Deltaretroviruses. *PLoS Comput Biol* 9(8):e1003162.
15. Hedges SB, Dudley J, & Kumar S (2006) TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22(23):2971-2972.
16. Lack JB, Roehrs ZP, Stanley CE, Ruedi M, & Van Den Bussche RA (2010) Molecular phylogenetics of Myotis indicate familial-level divergence for the genus Cistugo (Chiroptera). *Journal of Mammalogy* 91:976-992.
17. Miller-Butterworth CM, *et al.* (2007) A family matter: conclusive resolution of the taxonomic position of the long-fingered bats, miniopterus. *Mol Biol Evol* 24(7):1553-1561.
18. Johnson WE & Coffin JM (1999) Constructing primate phylogenies from ancient retrovirus sequences. *Proc Natl Acad Sci U S A* 96(18):10254-10260.
19. Kijima TE & Innan H (2010) On the estimation of the insertion time of LTR retrotransposable elements. *Mol Biol Evol* 27(4):896-904.
20. Zhuo X & Feschotte C (2015) Cross-Species Transmission and Differential Fate of an Endogenous Retrovirus in Three Mammal Lineages. *PLoS Pathog* 11(11):e1005279.
21. Kumar S & Subramanian S (2002) Mutation rates in mammalian genomes. *Proc Natl Acad Sci U S A* 99(2):803-808.
22. Perelman P, *et al.* (2011) A molecular phylogeny of living primates. *PLoS Genet* 7(3):e1001342.
23. Katzourakis A & Gifford RJ (2010) Endogenous viral elements in animal genomes. *PLoS Genet* 6(11):e1001191.
24. Fang J, Wang X, Mu S, Zhang S, & Dong D (2015) BGD: a database of bat genomes. *PLoS One* 10(6):e0131296.
25. Switzer WM, *et al.* (2009) Ancient, independent evolution and distinct molecular features of the novel human T-lymphotropic virus type 4. *Retrovirology* 6:9.
26. Miller-Butterworth CM, Jacobs DS, & Harley EH (2003) Strong population substructure is correlated with morphology and ecology in a migratory bat. *Nature* 424(6945):187-191.
27. Miller-Butterworth CM, Eick G, Jacobs DS, Schoeman MC, & Harley EH (2005) Genetic and phenotypic differences between south African long-fingered bats, with a global miniopterine phylogeny. *Journal of Mammalogy* 86:1121-1135.
28. Johnson M, *et al.* (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res* 36(Web Server issue):W5-9.
29. Smith MF & Patton JL (1991) Variation in mitochondrial cytochrome b sequence in natural populations of South American akodontine rodents (Muridae: Sigmodontinae). *Mol Biol Evol* 8(1):85-103.
30. Stadelmann B, Lin LK, Kunz TH, & Ruedi M (2007) Molecular phylogeny of New World Myotis (Chiroptera, Vespertilionidae) inferred from mitochondrial and nuclear DNA genes. *Mol Phylogenet Evol* 43(1):32-48.
31. Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772-780.
32. Guindon S, *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307-321.
33. Rambaut A & Grassly NC (1997) Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput Appl Biosci* 13(3):235-238.

Virology 485 (2015) 96–103

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Induction and characterization of a replication competent cervid endogenous gammaretrovirus (CrERV) from mule deer cells

Helena Fábryová [1,a], Tomáš Hron [a,1], Hana Kabíčková [b], Mary Poss [c], Daniel Elleder [a,*]

[a] Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, 14220 Prague, Czech Republic
[b] Military Health Institute, Department of Microbiology and Biological Research, 16001 Prague, Czech Republic
[c] Department of Biology, Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA 16801, USA

ARTICLE INFO

Article history:
Received 9 April 2015
Returned to author for revisions
15 May 2015
Accepted 4 July 2015
Available online 25 July 2015

Keywords:
Endogenous retrovirus
Xenotropism
Retrovirus interference

ABSTRACT

Endogenous retroviruses (ERVs) were acquired during evolution of their host organisms after infection and mendelian inheritance in the germline by their exogenous counterparts. The ERVs can spread in the host genome and in some cases they affect the host phenotype. The cervid endogenous gammaretrovirus (CrERV) is one of only a few well-defined examples of evolutionarily recent invasion of mammalian genome by retroviruses. Thousands of insertionally polymorphic CrERV integration sites have been detected in wild ranging mule deer (*Odocoileus hemionus*) host populations. Here, we describe for the first time induction of replication competent CrERV by cocultivation of deer and human cells. We characterize the physical properties and tropism of the induced virus. The genomic sequence of the induced virus is phylogenetically related to the evolutionarily young endogenous CrERVs described so far. We also describe the level of replication block of CrERV on deer cells and its capacity to establish superinfection interference.

Introduction

Endogenous retrovirus sequences constitute an integral part of all vertebrate genomes. They are generated following infection of the germline lineage of the host by an exogenous retrovirus and subsequent vertical inheritance of the integrated provirus form (Feschotte and Gilbert, 2012). The ERVs are classified into a large number of groups, whose diversity exceeds the currently circulating retrovirus species (Blomberg et al., 2009; Hayward et al., 2015). After the initial establishment of an integrated virus copy, which serves as a founder for a specific ERV group, further amplification and creation of new copies is enabled either by reinfection or by intracellular retrotransposition in the germline (Dewannieux et al., 2004; Jern and Coffin, 2008; Kanda et al., 2013).

Uncontrolled proliferation of ERVs in the genome would cause a burden for the host through mutagenic and various other effects. Therefore, there are multiple mechanisms that keep ERV expression and replication under control, most notably by transcriptional silencing (Liu et al., 2014; Rowe and Trono, 2011; Turelli et al., 2014). On the other hand, ERVs can be utilized for protection of the host from infecting retroviruses, a concept dubbed as “fighting fire with fire” (Malfavon-Borja and Feschotte, 2015). There are several well-documented cases in chickens, mice, cats and sheep, where endogenous envelope (Env) proteins can prevent the cell surface receptors from interacting with incoming retrovirus, resulting in a block of cellular entry (Malfavon-Borja and Feschotte, 2015). ERV-encoded proteins can also cause inhibition at several post-entry stages of infection (Arnaud et al., 2007b; Best et al., 1996; Monde et al., 2012). Another important way how ERVs can influence the outcome of retroviral infection is through recombination. ERV genomes can recombine among different endogenous loci or with related exogenous retroviruses. This can lead to the generation of fully infectious virus from two defective ERV genomes, or to the altered properties, for example altered tropism, of the exogenous partner involved in the recombination (Anai et al., 2012; Levy, 2008; Paprotka et al., 2011; Shimode et al., 2015; Young et al., 2012). In addition, through recombination with cellular genes, ERVs can form acutely transforming retroviruses (Kozak, 2015).

A practical classification is to consider ERVs as either “ancient” or “modern”, based on the time when they infiltrated the host genome (Armezzani et al., 2014). Most ERVs belong to the ancient category, where the genome invasion occurred long time ago in the evolutionary history of the host species, usually before the last speciation. Consequently, the individual ERV integrations are fixed in the host population, or even shared in phylogenetically related species. Modern ERVs entered the host genome more recently, mostly after speciation. Such ERV integrations have typically not yet reached fixation or been lost from the host lineage. At that stage, they are present in some individuals and absent in others,

* Corresponding author.
E-mail address: daniel.elleder@img.cas.cz (D. Elleder).
[1] Authors contributed equally.

http://dx.doi.org/10.1016/j.virol.2015.07.003
0042-6822/

which is denoted as insertional polymorphism. There are only a few well-studied examples of modern ERVs, these include the koala retrovirus (KoRV), endogenous Jaagsiekte sheep retroviruses (enJSRVs), porcine endogenous retroviruses (PERVs), endogenous feline leukemia viruses (enFeLVs) and other feline endogenous retroviruses (ERV-DCs), various mouse ERVs, and cervid endogenous gammaretrovirus (CrERV). The research on these viruses has lead to important insights into the process of genome invasion by an ERV and of the changes that accompany endogenization (Anai et al., 2012; Arnaud et al., 2007a; Lavillette and Kabat, 2004; Li et al., 2012; Oliveira et al., 2007; Tarlinton et al., 2006). In koalas, sheep and cats and mice, exogenous counterparts of the respective ERVs are circulating in natural populations and are associated with disease (Armezzani et al., 2014; Kozak, 2015; Levy, 2008; Xu et al., 2015). Replication-competent variants of PERV have also been reported (Preuss et al., 2006). In our previous studies we have advanced knowledge of the CrERV-mule deer model (Bao et al., 2014; Elleder et al., 2012; Kamath et al., 2014; Wittekindt et al., 2010). We have described an extensive collection of thousands of polymorphic endogenous retrovirus integration sites, comprehensively documenting the recent invasion of mule deer genomes by CrERV. The integration site patterns in individual deer were analyzed and revealed fine population structure and history of wild mule deer populations, with better resolution than in a parallel analysis performed with microsatellite markers (Kamath et al., 2014). However, all our previous work was focused on analysis of the integrated CrERV DNA or of the viral RNA expression. We have never obtained conclusive evidence of virus production or replication. Therefore, we attempted to replicate a previously published experiment (Aaronson et al., 1976), where primary blacktail deer (*O. hemionus colubmianus*, a subspecies of mule deer) cells were cocultured with human cell line. This led to the induction of replication competent gammaretrovirus species of hitherto unknown sequence, denoted deer kidney virus (DKV) (Aaronson et al., 1976; Barbacid et al., 1980).

In this study we report a successful induction of replication-competent CrERV from coculture of deer cells with a susceptible human cell line. We have characterized the physical properties of the induced virus, its phylogenetic relatedness to known endogenous CrERV copies, and its infectivity on deer and human cells. We also analyzed the capacity of the induced virus to establish interference to superinfection.

Results

CrERV can be induced by coculture of deer and human cells

Black-tailed deer primary kidney cells (OHK) and a human rhabdomyosarcoma cell line A673 were used in the coculture experiment. These cells were the same as those used in the original protocol (Aaronson et al., 1976). After approximately 30 days, RT activity in culture medium could be detected by a sensitive product-enhanced RT (PERT) assay (Fig. 1). At this point we stopped adding the OHK cells, which had served as a source of the induced virus. Due to much faster growth, only the human cells remained presumably in the subsequent continuation of the coculture. The RT activity continued to increase and eventually reached a plateau. The resulting RT level was still very low, approximately a thousand times lower than the values obtained for another endogenous gammaretrovirus, PERV (porcine endogenous retrovirus). The human and deer cells cultured separately tested negative in the PERT assay (data not shown). To confirm the identity of the induced virus species, viral cDNA was prepared from ultracentrifugation-concentrated culture fluids. Sequences highly identical to CrERV were obtained (full sequence of induced CrERV [CrERV-IND] is reported below). Therefore it is highly probable that DKV described by Aaronson, from which no sequence data is available, is identical to our recently reported CrERV.

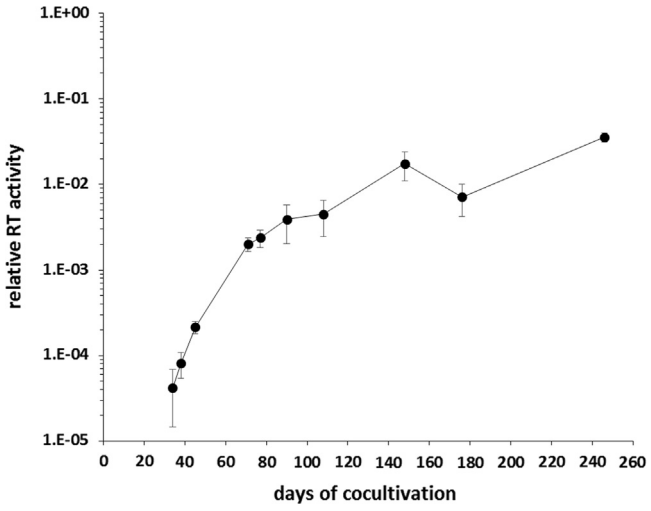

Fig. 1. Induction of virus from *O. hemionus* primary kidney cells cocultivated with human rhabdomyosarcoma cell line. At the indicated times, the RT activity in the culture supernatants was measured by a sensitive PERT assay. The results are expressed as means and standard deviations from triplicate assays.

CrERV particles

Next we examined whether the RT activity obtained from the coculture experiment belonged to particles of expected retrovirus buoyant density. The pelleted CrERV-IND was separated on iodixanol gradient and individual gradient fractions were tested by the PERT assay. For comparison we used virus particles of well-described endogenous gammaretrovirus, PERV (Bartosch et al., 2004). The RT activity peaked around the expected density of 1.1 g/ml, typical for retrovirus particles (Contreras-Galindo et al., 2012) (Fig 2A). We also obtained electron micrographs of both CrERV-IND and PERV (Fig. 2B).

The induced CrERV is infectious and xenotropic

Then, we evaluated the infectivity of the CrERV-IND particles. The virus inoculum from the coculture was used to infect naïve human and deer cells, and infectivity was assessed by the appearance of RT activity in the culture medium. Both human A673 and HEK 293 T cells could be infected (Fig. 4), however no RT activity was detected upon infection of deer OHK cells (data not shown). This is consistent with xenotropic characteristics of the induced virus, and is in agreement with observations reported by Aaronson et al. (1976) for DKV.

Relationship of induced CrERV sequence with known endogenous proviruses and construction of infectious molecular clone

To obtain the full sequence of CrERV-IND, we have amplified the provirus DNA from CrERV-infected HEK 293 T cells by long-range PCR. This sequence (deposited in Genbank under accession number KP261824) was 9,027 nucleotides long and had high identity across the entire length with a previously reported complete provirus genome, denoted CrERV-in7 (Elleder et al., 2012). There were intact open reading frames for all viral genes, *gag*, *pro/pol*, and *env* (Fig. 3B). We performed phylogenetic comparison of CrERV-IND with a set of previously published twelve endogenous CrERVs (Kamath et al., 2014). Because the full sequences of these CrERVs are not known, we have used an alignment of approximately 1.1 kb region in the 3'end of the virus genome to create the phylogeny. This region was identified

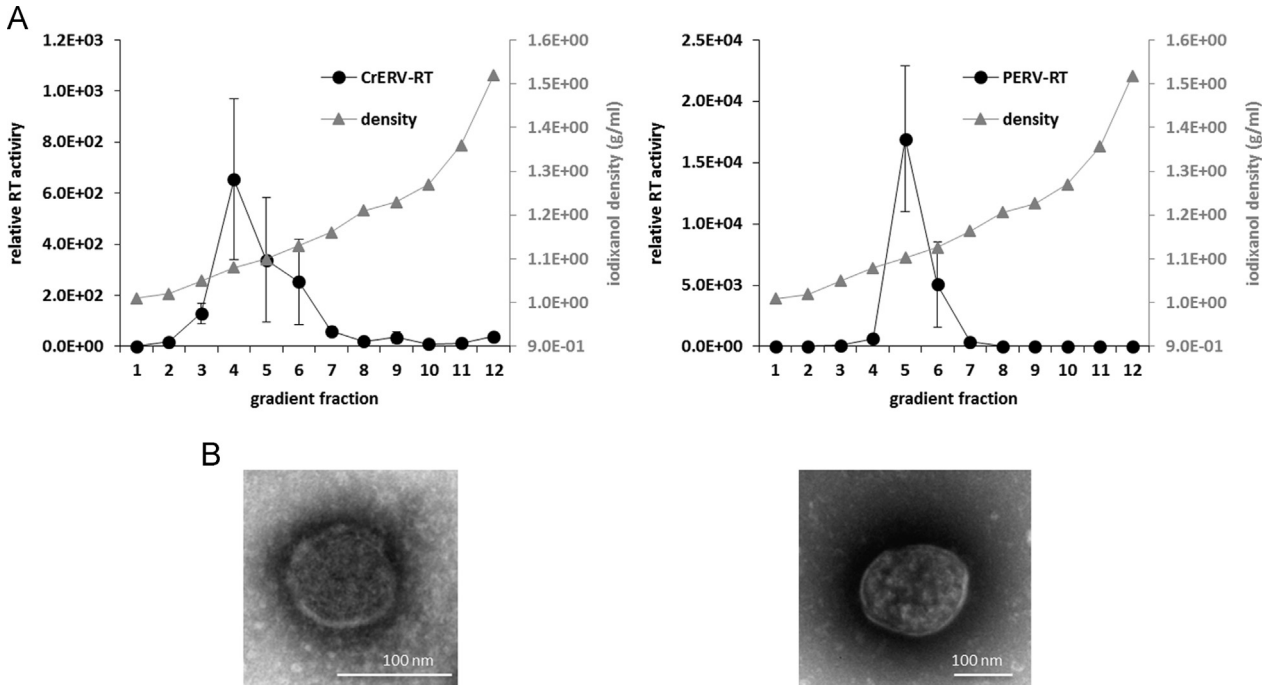

Fig. 2. Characterization of CrERV particles. (A) Pellets containing ultracentrifugation-concentrated virus particles were resuspended in PBS and layered on top of 10–50% stepwise iodixanol gradients. Following 17 h of centrifugation at 209,000 g, aliquots from each fraction were tested for RT activity by the PERT assay. Averages and standard deviations from triplicate assays are shown as black circles. Density of individual gradient fractions is depicted by gray triangles, with values displayed on the right axis. (B) Electron micrographs of retrovirus particles obtained by negative staining are shown, with 100 nm scale bar highlighted. In both A and B, CrERV is positioned on the left and PERV on the right.

previously (Kamath et al., 2014) to be convenient for phylogenetic analysis because it minimizes the influence of recombinant sequences. The CrERV-IND sequence clustered with good support with the large group of insertionally polymorphic, evolutionarily young CrERVs (Fig. 3 A). The CrERV-in1, which represents an evolutionarily old integration, was located on a separate branch.

We further compared the CrERV-IND sequence with four CrERVs for which we have full genomes available. Consistent with the results from the phylogenetic tree, the closest sequence to CrERV-IND was CrERV-in8, with only 24 genetic changes scattered over the entire 9- kb genome (Fig. 3B). CrERV-in4 and CrERV-in7 had 26 and 41 changes, respectively, relative to the CrERV-IND, and CrERV-in5 was even more distant. The ratio of nonsynonymous to synonymous substitutions in the coding regions was always higher in *env* than in *gag* and *pro/pol* (Fig. 3B). This is consistent with a higher degree of purifying selection in *gag* and *pro/pol* than in the *env* gene.

To standardize the work with the CrERV-IND, a molecular clone was constructed by subcloning the long range PCR product into pGEM-T Easy plasmid vector. The resulting construct pCrERV-IND was replication competent following transfection into HEK 293 T cells and subsequent infection of naïve cells. All subsequent analyses were performed with virus derived from this molecular clone.

CrERV infection kinetics in human cells

The RT activity of the CrERV-IND was very low even on permissive human cells. This is caused at least in part by very low titer of the virus. In end-point dilution experiments, the CrERV infectious titer on HEK 293 T cells was 10^{2-3} per ml. In addition, we tried to evaluate the infection kinetics using quantitative PCR methods. PERV was used again for comparison, and both viral inocula were normalized for RT activity. We have used both standard SYBR green real-time quantitative PCR and a highly sensitive digital droplet PCR method. The assay was detecting the newly formed virus DNA in the *env* gene region, corresponding to intermediate products of reverse transcription. The products could be detected several hours after infection of human cells and peaked around 8–24 h p.i., as described for other retroviruses (Mohammadi et al., 2013) (Fig. 4, left). The levels of newly made CrERV DNA were very low, about hundred times lower than for PERV. Based on the peak values being below $1x10^{-3}$ CrERV DNA copies per diploid cell genome, we estimate that less that 1 in 1000 cells was infected. However, during long-term culture of infected HEK 293 T cells, presumably all cells become infected in the course of about 3 weeks, because the virus DNA level reaches a plateau with values close to one copy per diploid cell genome (Fig. 4, right).

Early block of CrERV infection on deer cells, caused presumably by receptor interference

The induced CrERV was unable to productively infect deer cells and we wanted to characterize the level of the putative replication block. PCR-based assays used for human cells were not applicable because of the large background of hundreds of endogenous CrERV copies (Elleder et al., 2012) that would co-amplify together with the newly formed virus DNA in deer cells. Other relevant approaches, namely CrERV ENV pseudotypes of MLV and marker rescue assays were technically unsuccessful even after repeated attempts. Therefore, we used an alternative approach and by in vitro mutagenesis of the CrERV-IND molecular clone, we generated a sequence variant CrERV-mut. This construct contained several silent mutations in the *pol* gene (Fig. 5A). We have selected nucleotide variants not present in any of the previously described CrERV endogenous copies (Elleder et al., 2012; Kamath et al., 2014). This allowed us to design PCR primers that amplified only the newly generated CrERVmut DNA and not any of the endogenous CrERV copies or the parental CrERV virus (Fig. 5B, lanes 7 and 8). CrERV-mut virus was infectious, because it could infect naïve HEK-293 T cells (Fig. 5B, lanes 5 and 6). However, CrERV-mut newly made viral DNA was not detectable upon infection of deer cells (Fig. 5B, lanes 1 and 2). Therefore, either

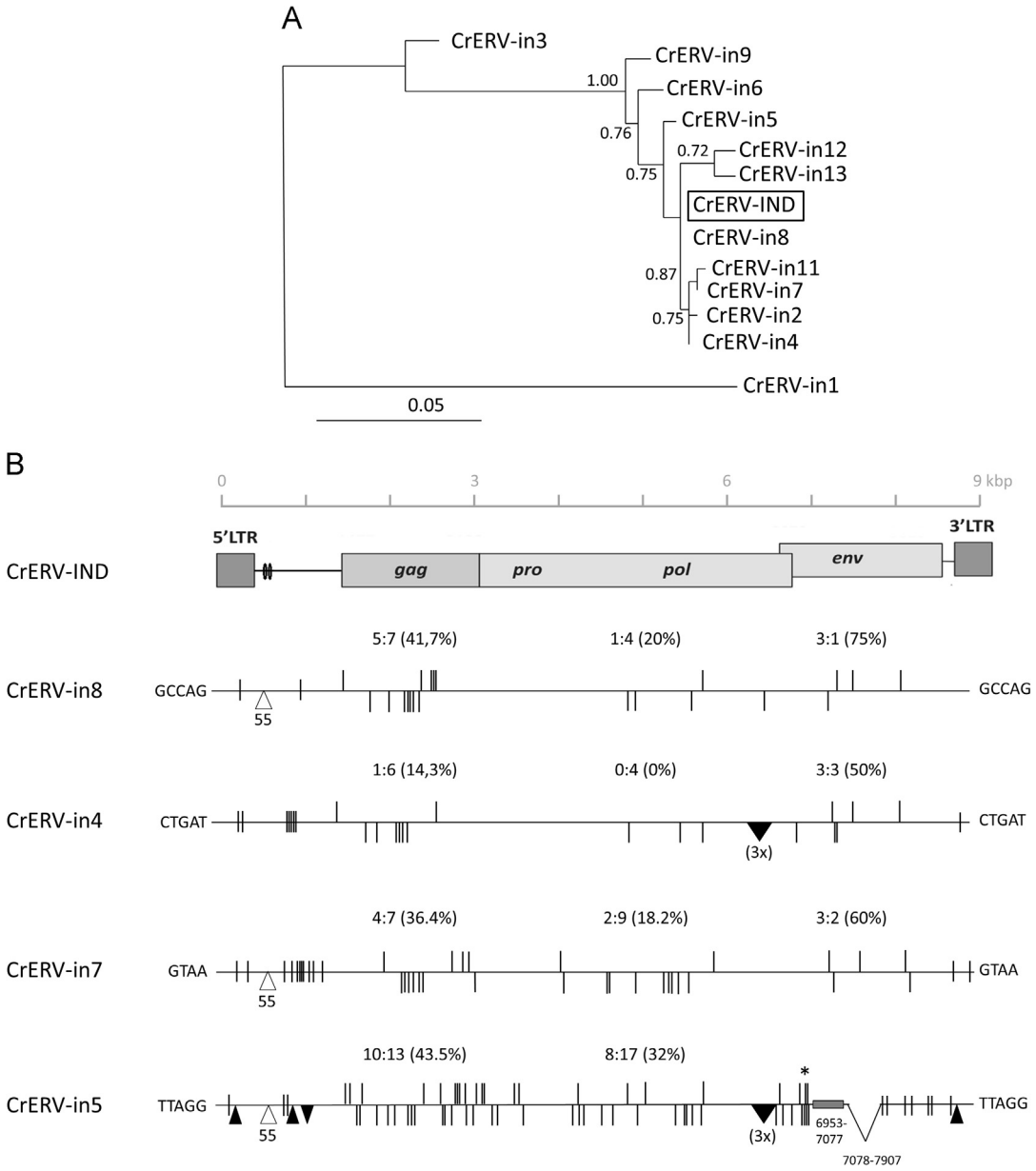

Fig. 3. Relationships between CrERV-IND sequence and endogenous CrERVs. (A) Phylogenetic tree generated from alignment of 1.1 kb region in the 3′end of the CrERV-IND and endogenous CrERVs. CrERV-IND position is highlighted by an open box. Bootstrap supports higher than 0.7 are indicated at the tree nodes. The branch length is proportional to the number of substitutions per site, with scale bar shown below the tree. (B) Comparison of the CrERV-IND sequence with four full endogenous CrERV genomes. The CrERV-IND genome is shown as a schematic at the top. Virus genes and LTRs are shown as shaded boxes and two direct 55- bp repeats in the leader region are displayed as small ovals. Every genome is depicted by horizontal black line, flanked with short target site duplication (TSD) sequences that arise during retrovirus integration. Short vertical lines indicate single nucleotide mismatches relative to the CrERV-IND sequence. Mismatches inside the coding regions for *gag*, *pro/pol* and *env* are shown above and below the horizontal line for the changes that result in nonsynonymous and synonymous substitutions, respectively, in the corresponding virus ORF. The numbers above each line represent the number of nonsynonymous: number of synonymous substitutions in each gene. The percentage of nonsynonymous mutations out of total mutations is shown in parentheses. Upward pointing black triangles indicate single nucleotide insertions, downward pointing triangles are single nucleotide deletions (the sign 3x denotes three closely spaced deletions). Open triangles represent insertion of third 55- bp direct repeat in the leader region. Premature stop codons are indicated by asterisk. Gray shaded box shows a region of CrERV-in5 *env* gene that is unalignable with CrERV-IND *env*. V-shaped line indicate region with large deletion.

receptor-mediated entry or some of the earliest steps of CrERV replication (virus uncoating, RT initiation), that precede virus DNA synthesis, are deficient in deer cells. To further test if receptor interference can block CrERV infection, we used the human cells chronically infected with CrERV. These cells presumably all harbor integrated CrERV (Fig. 4B) and express the virus envelope and therefore have the capacity to block cellular receptors used for virus entry. The chronically infected cells did not support generation of CrERVmut virus DNA (Fig. 5B, lanes 3 and 4) This is consistent with receptor interference being the cause of the resistance to CrERV on both deer cells and chronically infected human cells.

Discussion

Here we report the induction and characterization of a replication competent endogenous gammaretrovirus, CrERV, from mule deer cells in a coculture experiment. The process that led to the induction of CrERV production from the coculture is not known. It could have involved low production of infectious virus particles from deer cells, which then infected the permissive human cells. We have not detected virus production from deer cells by the PERT assay. However, this production could have been extremely low or intermittent. Among other potential mechanisms are rare spontaneous cell fusion

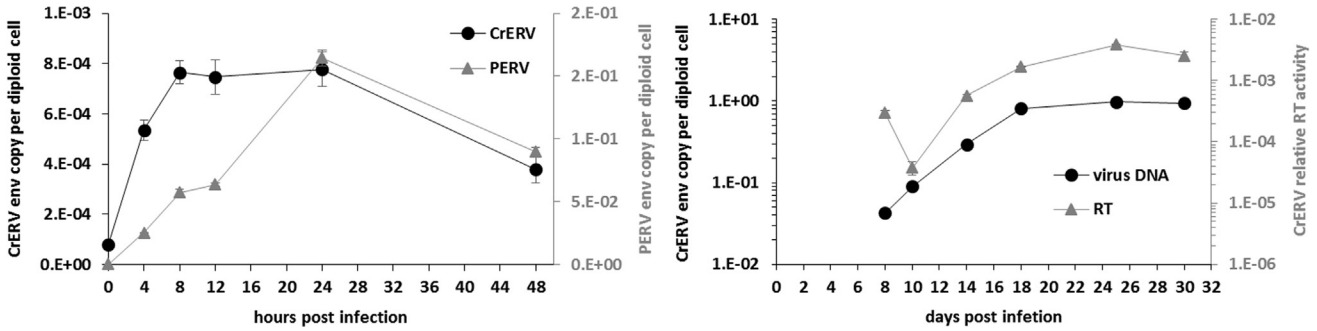

Fig. 4. CrERV infection kinetics on HEK 293 T cells. (Left) Quantification of newly made virus DNA in the first two days post infection. At indicated times, infected cells were collected and cellular lysates were prepared as described in Methods. Real-time PCR assays using MESA green and primers complementary to the virus *env* were used to quantify virus DNA levels. CrERV is depicted by black circles and PERV by gray triangles. (Right) Long-term infection of HEK 293 T cells by CrERV. Virus DNA levels in cellular lysates were quantified by digital droplet PCR method, using the same *env*-based primers as in the experiment depicted on the left diagram. The resulting values are shown as black circles. RT activity was determined from culture supernatants in parallel and is shown as gray triangles. The infection experiments were performed at least twice with similar results. Values from one representative experiment are shown, with averages and standard deviations from quantitative assays performed in triplicate.

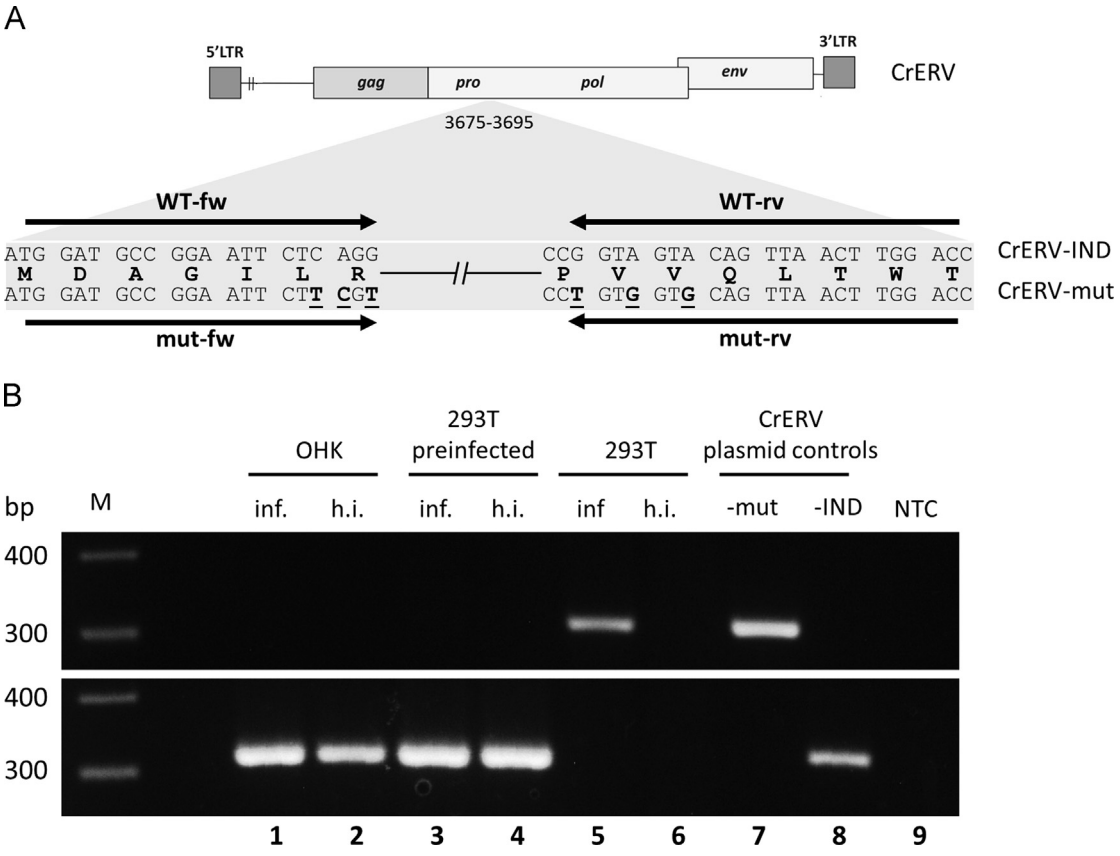

Fig. 5. Infection of human and deer cells by a CrERV sequence variant generated by in vitro mutagenesis, CrERV-mut. (A) A schematic depicting position and sequence of the primer pairs that specifically detect only CrERV-IND or CrERV-mut DNA. The upper and lower sequences represent the CrERV-IND and CrERV-mut genome, respectively. The nucleotide substitutions generated in CrERV-mut are underlined and bold. The arrows indicate positions of primers. Primers WT-fw and WT-rv amplify only the CrERV-IND virus DNA. Primers mut-fw and mut-rv amplify specifically the CrERV-mut DNA. (B) CrERV-mut was used to infect deer OHK cells (lane 1), HEK 293 T cells (lane 5) and HEK 293 T cells chronically infected with CrERV-IND (lane 3). Heat-inactivated (h.i.) virus was used in each case as a negative control to exclude virus DNA contamination (lanes 2, 4, and 6). Cells were harvested 20 h after infection and cellular lysates were prepared as described in Methods. CrERV-mut (lane 7) and CrERV-IND (lane 8) plasmid DNA was used as a control for specificity of PCR amplification. The upper panel shows PCR products generated with primers mut-fw and mut-rv, which detect specifically the CrERV-mut DNA. The lower panel shows PCR products generated with primers WT-fw and WT-rv. These primers amplify the "wild-type" variants of CrERV, i.e. the endogenous CrERVs in deer cells (lanes 1 and 2), and CrERV-IND in chronically infected 293 T cells (lanes 3 and 4). The experiments were performed twice with identical results; one representative experiment is displayed. M, molecular size marker; NTC, non-template control.

events, direct cell-to-cell transmission through virological synapses, or even transfer via exosomes (Sattentau, 2010; Wurdinger et al., 2012). In general, virus transmission or rescue by cocultivation was proven to be the most effective method for both endogenous and exogenous retroviruses (Agosto et al., 2015; Patience et al., 1997; Svoboda et al., 1963; Xu et al., 2013).

Several lines of evidence show that the induced CrERV is identical to the previously reported DKV, although the DKV genome has never been sequenced by the Aaronson laboratory. First, we tried to reproduce the setting of the original experiment, most importantly we used the same cell types in the coculture. Second, immunological methods established that DKV belonged to gammaretroviruses (Aaronson et al., 1976; Barbacid et al., 1980), as does CrERV. Third, DNA hybridization techniques using DKV probes established the presence of closely related endogenous viruses in all tested cervid species (Tronick et al., 1977). More

distantly related endogenous retroviruses were detected in other artiodactyls, e.g. in sheep. This pattern of distribution is consistent with the detection of various CrERV-related sequences in mammalian genomes (Elleder et al., 2012). A last argument pointing to the identity between CrERV and DKV is the similar tropism of both viruses, discussed below. DKV was shown to replicate on human cells but not deer cells, same as CrERV. In addition, DKV had a narrow xenotropic host range; it could replicate on horse cells, but not in cells of several additional mammalian species (Aaronson et al., 1976).

The genomic sequence of CrERV-IND did not show any apparent defect; it had complete open reading frames for all viral genes. In addition, the construction of an infectious molecular clone excluded the theoretical possibility that a helper virus from deer cells was needed for the replication of CrERV-IND. The obtained sequence was closest to the group of evolutionarily young and highly insertionally polymorphic CrERVs. This is consistent with the assumption that the youngest ERVs retain the highest capacity to be mobilized to form infectious progeny. Indeed, the closest relative to CrERV-IND was CrERV-in8, which was extremely rare in the population study, with only 1 positive animal out of 262 total (Kamath et al., 2014). However, none of the endogenous CrERVs was completely identical to CrERV-IND, therefore we cannot determine the source element that gave rise to the induced virus. There are hundreds of endogenous CrERVs in each deer genome; moreover, the deer cells used in the coculture were different from those we analyzed previously. We cannot exclude that recombination events between several endogenous CrERVs were involved in the generation of CrERV-IND, similar scenario was described for several other ERVs (Anai et al., 2012; Levy, 2008; Shimode et al., 2015; Young et al., 2012). Interestingly, although we analyzed only a few CrERV genomes, there seems to be a trend towards higher degree of purifying selection in *gag* and *pro/pol* than in env. Loss of *env* has been identified as a factor determining greater expansion of ERVs within the genome (Magiorkinis et al., 2012).

The buoyant density of CrERV-IND and electron micrographs of virus particles did not show obvious aberrations. However, the infectivity of CrERV-IND was very low even on permissive human cells. This could be due either to inherent defect of the CrERV-IND genome or to low permissivity of the human cells. The early kinetics of virus DNA production seemed normal and the virus eventually infected presumably all cells in the culture.

To explain the xenotropic nature of CrERV-IND, we tried to determine the level of the replication block on deer cells. Using the genetically modified variant CrERV-mut, we have shown that the block occurred before the viral DNA synthesis. The most plausible explanation is the receptor interference caused by the expression of endogenous *env* genes. To support this explanation, we have further shown that CrERV has the capacity to establish superinfection resistance on chronically infected human cells. However, because the cellular receptor for CrERV is not known, we cannot exclude that it is mutated in deer. Selection for such mutations in receptors for ERVs has been documented in endogenous avian leukosis viruses (ALV) and murine leukemia viruses (MLVs) (Barnard et al., 2006; Kozak, 2015). Treatment with tunicamycin, inhibitor of N-linked glycosylation, has been shown in some cases to abrogate resistance to retrovirus entry, by deglycosylation of the cellular receptors or the virus envelope proteins (Koo et al., 1994; Miller and Miller, 1992). However, tunicamycin treatment did not rescue sensitivity to CrERV upon infection of either deer cells or the chronically infected human cells (data not shown).

There is an intriguing paradox between the xenotropic nature of CrERV and its high efficiency of generation of new germline integrations in mule deer lineage in recent past. One possible solution to this paradox is that endogenous CrERVs are not expressed in the germline cells and therefore do not block the entry receptors. Alternatively, more variants of CrERV may exist that differ in receptor usage and can overcome the interference blocks, a mechanism described in FeLV and KoRV (Overbaugh et al., 2001; Xu et al., 2015). Even more complex mechanism was described in PERVs, where disruption of a highly conserved PHQ motif in the envelope glycoprotein enables transactivation of such viruses by unrelated gammaretroviral envelopes (Lavillette and Kabat, 2004). The PERVs with disrupted PHQ motif gain the ability to infect cells that lack the cognate PERV receptors and also to overcome restrictions caused by receptor interference. This property was suggested to provide novel opportunities to infect germ cells (Lavillette and Kabat, 2004). Interestingly, we observe a tendency toward disruption of the PHQ motif in the evolutionarily young CrERVs (D.E., personal communication). The analysis of the possible underlying mechanisms is under way in our laboratory.

Materials and methods

Cells and co-cultivation protocol

Human rhabdomyosarcoma cell line A-673 (ATCC product number CRL-1598) and primary *O. hemionus* kidney cells (OHK, ATCC product number CRL-6193) were grown in Dulbecco's modified Eagle's medium (Sigma-Aldrich, St. Louis, MO) supplemented with 10% fetal calf serum, penicillin and streptomycin. HEK 293 T cells were grown in the same conditions, with serum supplements changed to 4% fetal calf serum and 4% calf serum. HEK-293 T cells producing PERV 14/220 (Bartosch et al., 2004) were used as a source of porcine gammaretroviral particles. All cells were cultured in a humidified incubator at 37 °C and 5% CO_2. The co-cultivation experiment was started by mixing equal numbers of deer CRL-6193 and human A-673 cells. Every week, fresh cells from both species were added to the coculture at 1:1 ratio. At indicated time points, samples of the culture fluids were harvested for the RT assay. The samples were spun at 3000 RPM for 5 min to remove cell debris, filtered by a 0.22 μm syringe filter and frozen at −80 °C before further use.

PERT assay

The PERT assay was adapted from previously published protocols (Lovatt et al., 1999; Pizzato et al., 2009; Sears and Khan, 2003). The samples (2 μl of culture supernatant or gradient fraction) were lysed in 8 μL of solution containing 1% TRITON X-100, 0.4 U/μL RNasin (Promega), and 1x ProtoscriptII buffer (New England Biolabs) at room temperature for 30 min. Then, two mastermixes (A and B) were prepared, with the following amounts per one reaction: Mix A contained 20 ng of the template RNA of MS2 phage (Roche), 0.5 ul of the reverse primer (5′- GCCTTAGCAGTGCCCTGTCT) and 10.1 ul water. Mix B contained 3.6 ul of 5x ProtoscriptII buffer, 2 ul of 100 mM DTT, 0.8 ul of 10 mM dNTP2, and 6.4 ul water. Mix A was incubated at 65 °C for 5 min and slowly cooled down to allow primer annealing. Next the mixes A and B were pooled and aliqoted by 18 ul. To each aliquot, 2 ul of the lysates were added and incubated at 37 °C for 30 min (reverse transcription step), then inactivated at 70 °C for 10 min, The newly generated MS2 cDNA was quantified by real-time PCR assay with forward (5′- AACATGCTCGAGGGCCTTA) and reverse primers and probe (FAM-TGGGATGCTCCTACATG-TAMRA). Each reaction contained 1.5 ul of the cDNA sample, 1xqPCR master mix (Eurogentec, Seraing, Belgium), 7.5 pmol of each primer and 3.75 pmol of probe in a total volume of 15 ul. The samples were run on a Bio-Rad CFX96™ Real-Time instrument with a three-step protocol: 1 cycle of 10 min at 95 °C and then 45 cycles consisting of 15 s at 95 °C, 20 s at 60 °C and 20 s at 72 °C. Cycles of quantification (Cq) values

were generated by the CFX Manager software. With each run, one calibrator sample was assayed and all values were expressed as relative values compared to the calibrator.

Iodixanol gradient

Iodixanol (OptiPrep™) was purchased from Axis Shield (Dundee, United Kingdom). Thirty milliliters of cell free supernatants from virus-producing cells were cleared from cellular debris by low-speed centrifugation (3000 RPM for 5 min at 4 °C) and then centrifuged through a 20% iodixanol cushion in a SW28 rotor (Beckman Coulter Pasadena, CA) for 2 h at 23,000 RPM. The centrifuged pellet was resuspended in 1 ml of PBS. Two-milliliter layers containing 50%, 40%, 30%, 20%, and 10% iodixanol were pipetted in tubes for the SW41Ti rotor (Beckman Coulter) and the resuspended pellet in PBS was applied on top. The gradient was centrifuged for 17 h at 35,000 rpm at 4 °C. Gradient fractions were collected from the top and their density was determined by refractometry. Aliquots from each fraction were used for the PERT assay.

Electron microscopy

Virus particles from the culture medium of infected cells were pelleted by ultracentrifugation as described above and fixed in 2% formaldehyde. Samples negatively stained with 3% phosphotungstic acid (PTA) were then viewed with Jeol JEM, 2000 CX microscope (JEOL, Arishima, Japan).

Plasmid construction

For the preparation of replication-competent CrERV DNA clone, two partially overlapping proviral fragments were separately amplified from infected HEK-293 T cells using PCR. Primers for amplification of 5′ CrERV fragment were: 5′–AACCGCGGCCGCTGTAGGGAGAACAAACGGAATGTAGAAAG-3′ (NotI restriction site used for cloning is underlined) and 5′–CAGGGGTAGGCTGAAAAAGGCATC-3′. Primers for amplification of 3′ CrERV fragment were: 5′–TACCCATATGTGTTATGCCCGATGTCCGAATCC-3′ (NdeI restriction site used for cloning is underlined) and 5′–GCCCTCAGAGGTCATAGCACCAGA-3′. Amplified DNA fragments covered the entire proviral sequence and contained unique EcoRI restriction site in the overlapping region. The DNA fragments were digested with NotI/EcoRI and NdeI/EcoRI for the 5′ and 3′ regions, respectively. Consequently, both cleaved fragments were sequentially ligated into the pGem-T Easy cloning vector (Promega, Madison, WI), creating intact CrERV proviral DNA clone (pCrERV-IND). To create the modified variant of CrERV (CrERV-mut), 298 bp-long region of viral *pol* gene was PCR-amplified using mutagenic primers which contain several mismatching bases and EcoRI/HpaI restriction sites naturally occurring in viral sequence: 5′–CCGGAATTCTTCGTCGGTGCCAA and 5′–CAAGTTAACTGCACCACAGGTTGG. Original sequence in pCrERV-IND plasmid was replaced by this mutated fragment using EcoRI and HpaI enzymes, which generated the pCrERV-mut construct.

Virus preparation and cell infection

HEK-293 T cells were transfected with pCrERV-IND or pCrERV-mut plasmids using X-tremeGENE HP DNA Transfection Reagent (Roche s.r.o., Prague, Czech Republic). Specifically, 5×10^4 cells were seeded onto 24-well plate and next day transfected by addition of 100 μl of culture medium containing 0.5 μg of DNA and 0.5 μl of the transfection reagent. Cells were grown for approximately 30 days until the viral DNA copy number reached plateau. Afterwards, 100 μl of filtered cell supernatant was used for infection of naïve HEK-293 T cells seeded on 24-well plate (5×10^4 per well) a day before. After 1 h of incubation at 37 °C, 400 μl of fresh medium was added, and the cells were further incubated in CO_2 at 37 °C. For determining the infectious titer of CrERV, 10-fold dilutions of virus were inoculated in triplicate wells. The cells were passaged for 4 weeks to allow virus spread, and scored for virus presence using both PCR assays and PERT assay.

PCR-based detection of viral DNA (qPCR, digital droplet PCR and standard analytical PCR)

For the determination of reversely-transcribed viral DNA, infected cells were harvested at appropriate time points, washed by PBS, and incubated in lysis buffer (10 mM Tris–HCl, pH 8.0, 1 mM EDTA, 0.2 mM $CaCl_2$, 0.001% Triton X-100, 0.001% SDS, 1 mg/ml proteinase K) for 60 min at 58 °C, followed by 10 min protease heat-inactivation at 95 °C. Primers for quantitative PCR (5′–TGACCCCATGTTTGAATGTG and 5′–GAGGACAGCTCCTTGGTTTG) were designed to conserved region of CrERV *env* gene using Primer3Plus software (Untergasser et al., 2007). For the quantification of PERV DNA, primers 5′–AGGTGGTGGGCATGTAATACTG and 5′–ACACTCGGGGAACAATTTGG, also situated in *env* gene, were used. MESA GREEN qPCR mastermix (Eurogentec, Seraing, Belgium) was used for standard real-time quantitative PCR. Each reaction mixture had a total volume of 20 μl, containing 2 μl of the cell lysate and 300 nM (each) the forward and reverse primers. The samples were run on a Bio-Rad CFX96™ Real-Time System (Bio-Rad, Hercules, CA) with a two-step protocol (1 cycle of 5 min at 95 °C and then 44 cycles consisting of 15 s at 95 °C and 60 s at 60 °C), followed by melting curve analysis in the CFX Manager software (Bio-Rad) to ensure the specificity of the amplification. An absolute standard curve for each assay was obtained by using as templates serial dilutions of a plasmid containing the corresponding amplicon. The results were normalized using the parallel amplification of a single-copy genomic locus in porphobilinogen deaminase gene (Konig et al., 2008).

For highly accurate absolute quantification of viral DNA, droplet digital PCR system QX200 (Bio-Rad, Hercules, CA) was used. Each reaction mixture had a total volume of 20 μl, containing 1x QX200 ddPCR Evagreen Supermix (Bio-Rad), 2 μl of the cell lysate (1–5 ng DNA), and 250 nM (each) the forward and reverse primers. The reactions were treated for droplet generation according to the manufacturer's manual and then amplified with the following conditions: 1 cycle of 5 min at 95 °C and then 40 cycles consisting of 15 s at 95 °C and 40 s at 59 °C followed by 1 cycle of 5 min at 72 °C, 5 min at 4 °C and 5 min at 90 °C. Samples were analyzed by droplet reader and QuantaSoft software (Bio-Rad) with thresholds set manually.

For specific detection of modified CrERV variant (CrERV-mut) in infected cells, PCR primers complementary to the modified region of viral DNA were designed: 5′–GGATGCCGGAATTCTTCGT and 5′–GTCCAAGTTAACTGCACCACA. Primers complementary to the wild-type CrERV sequence were used as a positive control: 5′–GGATGCCGGAATTCTCAGG-3′ and 5′–GGGTCCAAGTTAACTGTACTACC-3′. OneTaq mastermix (New England Biolabs, Ipswich, MA) was used for analytic PCR amplification. Each reaction mixture had a total volume of 20 μl, containing 2 μl of the cell lysate and 200 nM (each) the forward and reverse primers. The samples were run with a following protocol: 1 cycle of 3 min at 94 °C and then 32 cycles consisting of 20 s at 94 °C, 30 s at 60 °C and 30 s at 68 °C finished by 1 cycle of 5 min at 68 °C and analyzed by agarose gel electrophoresis.

Phylogenetic tree

The 3′ portion of 13 CrERV sequences (approximately 1100 bp) was aligned using Muscle (Edgar, 2004) and the phylogeny was generated with PhyML (Guindon et al., 2010) and the HKY85 model with a gamma distribution.

Acknowledgments

We would like to thank Petr Pajer, Jiří Hejnar and Jan Svoboda for helpful discussions. This work was supported by program NÁVRAT (LK11215) provided by the Czech Ministry of Education, Youth and Sports, and by USGS grant 06HQAG0131.

References

Aaronson, S.A., Tronick, S.R., Stephenson, J.R., 1976. Endogenous type C RNA virus of *Odocoileus hemionus*, a mammalian species of New World origin. Cell 9, 489–494.
Agosto, L.M., Uchil, P.D., Mothes, W., 2015. HIV cell-to-cell transmission: effects on pathogenesis and antiretroviral therapy. Trends Microbiol. 23, 289–295.
Anai, Y., Ochi, H., Watanabe, S., Nakagawa, S., Kawamura, M., Gojobori, T., Nishigaki, K., 2012. Infectious endogenous retroviruses in cats and emergence of recombinant viruses. J. Virol. 86, 8634–8644.
Armezzani, A., Varela, M., Spencer, T.E., Palmarini, M., Arnaud, F., 2014. "Menage a Trois": the evolutionary interplay between JSRV, enJSRVs and domestic sheep. Viruses 6, 4926–4945.
Arnaud, F., Caporale, M., Varela, M., Biek, R., Chessa, B., Alberti, A., Golder, M., Mura, M., Zhang, Y.P., Yu, L., Pereira, F., Demartini, J.C., Leymaster, K., Spencer, T.E., Palmarini, M., 2007a. A paradigm for virus-host coevolution: sequential counter-adaptations between endogenous and exogenous retroviruses. PLoS Pathog. 3, e170.
Arnaud, F., Murcia, P.R., Palmarini, M., 2007b. Mechanisms of late restriction induced by an endogenous retrovirus. J. Virol. 8111441–11451
Bao, L., Elleder, D., Malhotra, R., DeGiorgio, M., Maravegias, T., Horvath, L., Carrel, L., Gillin, C., Hron, T., Fábryová, H., Hunter, D., Poss, M., 2014. Computational and statistical analyses of insertional polymorphic endogenous retroviruses in a non-model organism. Computation 2, 221–245.
Barbacid, M., Daniel, M.D., Aaronson, S.A., 1980. Immunological relationships of OMC-1, an endogenous virus of owl monkeys, with mammalian and avian type C viruses. J. Virol. 33, 561–566.
Barnard, R.J., Elleder, D., Young, J.A., 2006. Avian sarcoma and leukosis virus-receptor interactions: from classical genetics to novel insights into virus-cell membrane fusion. Virology 344, 25–29.
Bartosch, B., Stefanidis, D., Myers, R., Weiss, R., Patience, C., Takeuchi, Y., 2004. Evidence and consequence of porcine endogenous retrovirus recombination. J. Virol. 78, 13880–13890.
Best, S., Le Tissier, P., Towers, G., Stoye, J.P., 1996. Positional cloning of the mouse retrovirus restriction gene Fv1. Nature 382, 826–829.
Blomberg, J., Benachenhou, F., Blikstad, V., Sperber, G., Mayer, J., 2009. Classification and nomenclature of endogenous retroviral sequences (ERVs): problems and recommendations. Gene 448, 115–123.
Contreras-Galindo, R., Kaplan, M.H., Contreras-Galindo, A.C., Gonzalez-Hernandez, M.J., Ferlenghi, I., Giusti, F., Lorenzo, E., Gitlin, S.D., Dosik, M.H., Yamamura, Y., Markovitz, D.M., 2012. Characterization of human endogenous retroviral elements in the blood of HIV-1-infected individuals. J. Virol. 86, 262–276.
Dewannieux, M., Dupressoir, A., Harper, F., Pierron, G., Heidmann, T., 2004. Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. Nature Genet. 36, 534–539.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Elleder, D., Kim, O., Padhi, A., Bankert, J.G., Simeonov, I., Schuster, S.C., Wittekindt, N.E., Motameny, S., Poss, M., 2012. Polymorphic integrations of an endogenous gammaretrovirus in the mule deer genome. J. Virol. 86, 2787–2796.
Feschotte, C., Gilbert, C., 2012. Endogenous viruses: insights into viral evolution and impact on host biology. Nat. Rev. Genet. 13, 283–296.
Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.
Hayward, A., Cornwallis, C.K., Jern, P., 2015. Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. Proc. Natl. Acad. Sci. USA 112, 464–469.
Jern, P., Coffin, J.M., 2008. Effects of retroviruses on host genome function. Ann. Rev. Genet. 42, 709–732.
Kamath, P.L., Elleder, D., Bao, L., Cross, P.C., Powell, J.H., Poss, M., 2014. The population history of endogenous retroviruses in mule deer (*Odocoileus hemionus*). J. Hered. 105, 173–187.
Kanda, R.K., Tristem, M., Coulson, T., 2013. Exploring the effects of immunity and life history on the dynamics of an endogenous retrovirus. Philos. Trans. R. Soc. Lond. Ser B Biol. Sci. 368, 20120505.
Konig, R., Zhou, Y., Elleder, D., Diamond, T.L., Bonamy, G.M., Irelan, J.T., Chiang, C.Y., Tu, B. P., De Jesus, P.D., Lilley, C.E., Seidel, S., Opaluch, A.M., Caldwell, J.S., Weitzman, M.D., Kuhen, K.L., Bandyopadhyay, S., Ideker, T., Orth, A.P., Miraglia, L.J., Bushman, F.D., Young, J.A., Chanda, S.K., 2008. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell 135, 49–60.
Koo, H.M., Parthasarathi, S., Ron, Y., Dougherty, J.P., 1994. Pseudotyped REV/SRV retroviruses reveal restrictions to infection and host range within members of the same receptor interference group. Virology 205, 345–351.
Kozak, C.A., 2015. Origins of the endogenous and infectious laboratory mouse gammaretroviruses. Viruses 7, 1–26.
Lavillette, D., Kabat, D., 2004. Porcine endogenous retroviruses infect cells lacking cognate receptors by an alternative pathway: implications for retrovirus evolution and xenotransplantation. J. Virol. 78, 8868–8877.
Levy, L.S., 2008. Advances in understanding molecular determinants in FeLV pathology. Vet. Immunol. Immunopathol. 123, 14–22.
Li, J., Akagi, K., Hu, Y., Trivett, A.L., Hlynialuk, C.J., Swing, D.A., Volfovsky, N., Morgan, T.C., Golubeva, Y., Stephens, R.M., Smith, D.E., Symer, D.E., 2012. Mouse endogenous retroviruses can trigger premature transcriptional termination at a distance. Genome Res. 22, 870–884.
Liu, S., Brind'Amour, J., Karimi, M.M., Shirane, K., Bogutz, A., Lefebvre, L., Sasaki, H., Shinkai, Y., Lorincz, M.C., 2014. Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. Genes Dev. 28, 2041–2055.
Lovatt, A., Black, J., Galbraith, D., Doherty, I., Moran, M.W., Shepherd, A.J., Griffen, A., Bailey, A., Wilson, N., Smith, K.T., 1999. High throughput detection of retrovirus-associated reverse transcriptase using an improved fluorescent product enhanced reverse transcriptase assay and its comparison to conventional detection methods. J. Virol. Methods 82, 185–200.
Magiorkinis, G., Gifford, R.J., Katzourakis, A., De Ranter, J., Belshaw, R., 2012. Env-less endogenous retroviruses are genomic superspreaders. Proc. Natl. Acad. Sci. USA 109, 7385–7390.
Malfavon-Borja, R., Feschotte, C., 2015. Fighting fire with fire: endogenous retrovirus envelopes as restriction factors. J. Virol. 89, 4047–4050.
Miller, D.G., Miller, A.D., 1992. Tunicamycin treatment of CHO cells abrogates multiple blocks to retrovirus infection, one of which is due to a secreted inhibitor. J. Virol. 66, 78–84.
Mohammadi, P., Desfarges, S., Bartha, I., Joos, B., Zangger, N., Munoz, M., Gunthard, H.F., Beerenwinkel, N., Telenti, A., Ciuffi, A., 2013. 24h in the life of HIV-1 in a T cell line. PLoS Pathog. 9, e1003161.
Monde, K., Contreras-Galindo, R., Kaplan, M.H., Markovitz, D.M., Ono, A., 2012. Human endogenous retrovirus K Gag coassembles with HIV-1 Gag and reduces the release efficiency and infectivity of HIV-1. J. Virol. 86, 11194–11208.
Oliveira, N.M., Satija, H., Kouwenhoven, I.A., Eiden, M.V., 2007. Changes in viral protein function that accompany retroviral endogenization. Proc. Natl. Acad. Sci. USA 104, 17506–17511.
Overbaugh, J., Miller, A.D., Eiden, M.V., 2001. Receptors and entry cofactors for retroviruses include single and multiple transmembrane-spanning proteins as well as newly described glycophosphatidylinositol-anchored and secreted proteins. Microbiol. Mol. Biol. Rev. 65, 371–389.
Paprotka, T., Delviks-Frankenberry, K.A., Cingoz, O., Martinez, A., Kung, H.J., Tepper, C.G., Hu, W.S., Fivash Jr., M.J., Coffin, J.M., Pathak, V.K., 2011. Recombinant origin of the retrovirus XMRV. Science 333, 97–101.
Patience, C., Takeuchi, Y., Weiss, R.A., 1997. Infection of human cells by an endogenous retrovirus of pigs. Nat. Med. 3, 282–286.
Pizzato, M., Erlwein, O., Bonsall, D., Kaye, S., Muir, D., McClure, M.O., 2009. A one-step SYBR Green I-based product-enhanced reverse transcriptase assay for the quantitation of retroviruses in cell culture supernatants. J. Virol. Methods 156, 1–7.
Preuss, T., Fischer, N., Boller, K., Tonjes, R.R., 2006. Isolation and characterization of an infectious replication-competent molecular clone of ecotropic porcine endogenous retrovirus class C. J. Virol. 80, 10258–10261.
Rowe, H.M., Trono, D., 2011. Dynamic control of endogenous retroviruses during development. Virology 411, 273–287.
Sattentau, Q.J., 2010. Cell-to-cell spread of retroviruses. Viruses 2, 1306–1321.
Sears, J.F., Khan, A.S., 2003. Single-tube fluorescent product-enhanced reverse transcriptase assay with Ampliwax (STF-PERT) for retrovirus quantitation. J. Virol. Methods 108, 139–142.
Shimode, S., Nakagawa, S., Miyazawa, T., 2015. Multiple invasions of an infectious retrovirus in cat genomes. Sci. Rep. 5, 8164.
Svoboda, J., Chyle, P., Simkovic, D., Hilgert, I., 1963. Demonstration of the absence of infectious Rous virus in rat tumour XC, whose structurally intact cells produce Rous sarcoma when transferred to chicks. Folia Biol. 9, 77–81.
Tarlinton, R.E., Meers, J., Young, P.R., 2006. Retroviral invasion of the koala genome. Nature 442, 79–81.
Tronick, S.R., Golub, M.M., Stephenson, J.R., Aaronson, S.A., 1977. Distribution and expression in mammals of genes ralated to an endogenous type C RNA virus of *Odocoileus hemionus*. J. Virol. 23, 1–9.
Turelli, P., Castro-Diaz, N., Marzetta, F., Kapopoulou, A., Raclot, C., Duc, J., Tieng, V., Quenneville, S., Trono, D., 2014. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome Res. 24, 1260–1270.
Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A., 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. 35, W71–74.
Wittekindt, N.E., Padhi, A., Schuster, S.C., Qi, J., Zhao, F., Tomsho, L.P., Kasson, L.R., Packard, M., Cross, P., Poss, M., 2010. Nodeomics: pathogen detection in vertebrate lymph nodes using meta-transcriptomics. PloS One 5, e13432.
Wurdinger, T., Gatson, N.N., Balaj, L., Kaur, B., Breakefield, X.O., Pegtel, D.M., 2012. Extracellular vesicles and their convergence with viral pathways. Adv. Virol. 2012, 767694.
Xu, W., Gorman, K., Santiago, J.C., Kluska, K., Eiden, M.V., 2015. Genetic diversity of koala retroviral envelopes. Viruses 7, 1258–1270.
Xu, W., Stadler, C.K., Gorman, K., Jensen, N., Kim, D., Zheng, H., Tang, S., Switzer, W. M., Pye, G.W., Eiden, M.V., 2013. An exogenous retrovirus isolated from koalas with malignant neoplasias in a US zoo. Proc. Natl. Acad. Sci. USA 110, 11547–11552.
Young, G.R., Eksmond, U., Salcedo, R., Alexopoulou, L., Stoye, J.P., Kassiotis, G., 2012. Resurrection of endogenous retroviruses in antibody-deficient mice. Nature 491, 774–778.