CHARLES UNIVERSITY

Faculty of Sciences

Molecular and Cell Biology, Genetics and Virology

Host-virus interactions of mammalian endogenous retroviruses

Interakce savčích endogennich retrovirů a jejich hostitelů

PhD thesis

Mgr. Helena Farkašová

Supervisor: MUDr. Daniel Elleder, PhD.

Laboratory of Viral and Cellular Genetics

Institute of Molecular Genetics of the ASCR

Prague, 2017

©Tina Bone, Roe deer and Finches

“And above all, watch with glittering eyes the whole world around you because the greatest secrets are always hidden in the most unlikely places. Those who don't believe in magic will never find it.”

Roald Dahl

©Lucie Straková, Issue cover proposal for Molecular Biology and Evolution

DECLARATION

I hereby certify that I have written this thesis independently and that I have not used other than the cited sources.This thesis has not been submitted for any other degree or purposes.

January 2017, Prague .....

Mgr. Helena Farkašová

*Many Figures were made by one and thorougly remade a couple times by three authors (me, Tomáš Hron, and Daniel Elleder); and many results were obtained by one and reproduced by the other two. On top of it, we did have our fingers in many pies (not all of them discussed in this thesis, for some were not proven to be as tasty in the running) for Daniel Elleder was forever bursting with new ideas and me and Tomáš were never too hesitant to jump into new experiments.

PREFACE

This thesis discusses four endogenous retroviruses. The first one is an endogenous Lentivirus detected in the genome of *Galeopterus variegatus*. The second is (rather a remnant of) an endogenous Deltaretrovirus described in the genomes of Miniopteridae bats. The third one is an endogenous Gammaretrovirus in the genome of *Odocoileus hemionus*; and the fourth is the presumed Gammaretrovirus present in the *Cricetulus griseus* genome (or rather in the cells obtained from this species and widely used in biotechnology).

These animals come from only remotely related taxa and the studied retroviruses come from several groups. Therefore, in the introduction, I try to discuss the general phenomenon of endogenization, the individual groups of retroviruses, and the host restriction towards retroviral infection. The overview of the possible outcomes of the presence of an endogenous retrovirus for the host, and other aspects of the presence of an endogenous retrovirus in the animal genome are discussed not so elaborately.

The Methods section is divided into sub-chapters, each discussing methods used in an individual project connected to a particular retrovirus. This is so that the reader has an easier job following this thesis. If a particular method is used in various projects, it is described only once and a cross-reference is included to it in the other project section.

The results and discussion sections normally occurring in two individual chapters are merged, for with so many different projects, I found it easier to follow for the reader. In many cases, the methods, results or discussion texts might highly resemble the ones occurring in the published papers. Several figures are the ones included in the attached manuscripts, for several results discussed in this thesis or closely connected to it have been already published. All of these manuscripts are in the Supplement to this thesis. I tried to include the data I feel I contributed with to the presented papers (if not stated otherwise further on in the text), even though it is complicated in some cases (due to the facts stated in the declaration).

I also tried to discuss the data not published in the papers (for they are either preliminary or were proven to be a dead end analyzes) more in detail than the ones that could be found in the attached manuscripts. These data and figures from them are generally rather preliminary, yet already informative.

Taken together, I believe the experiments reported in this thesis would add to our understanding of the assorted diversity of interactions between endogenous retroviruses and their hosts.

ACKNOWLEDGMENT

I would like to acknowledge:

- My supervisor MUDr. Daniel Elleder, PhD for an incredible amount of patience, work, time, and energy he devoted to the projects related to this thesis as well as to my tutoring (including daily consultations) and for thoroughly revising this thesis.
- My closest co-worker Mgr. Tomáš Hron for all of the help, hard work, and inspiring attittude. Gentlemen, this was a great journey indeed!
- Jiří Hejnar for running the Laboratory of Viral and Cellular Genetics at Institute of Molecular Genetics in the fashion that coming to work is always something to look forward to.
- Jan Pačes for helping with the bioinformatics issues and all the tea he made us during the hours we sat in his office.
- The whole team of the Laboratory of Viral and Cellular Genetics for help and technical advice. Especially Kateřina Trejbalová, Filip Šenigl, Dana Kučerová; and Josef Geryk.
- Cyrili Bařinka and Mass Spec core facility at BIOCEV for helping us with the protein assays related to CHOK1 cells project.
- Hana Kabíčková for helping us with the electron microscopy.
- Petr Pajer for occasionally sequencing the unsequencable and Vladimír Pečenka for PCR advice.
- Mary Poss for being a driving force behind the initial CrERV studies and performing the pattern PCRs.
- The whole team of the Laboratory of Viral and Cellular Genetics again, for sustaining a relaxed, yet motivational environment and being not only co-workers but also friends. Moreover, Tom, Dalibor, Katy, Martina, Kryštof, Anička, and Albert for being a loyal lunch & laugh company.
- For all the people at IMG working really hard to socialize, help and advise PhD students whenever possible; for organizing PhD boot camp; biennial progress reports; PhD conferences; organizing courses and afterparties.
- For all the core facility staff at IMG for all the help and consultations (especially with flow cytometry and cell sorting), medium preparation, maintaining cell storage in liquid nitrogen, preparing plates, etc.
- The foundation “Nadání, Josefa, Marie a Zdeňky Hlávkových” for providing me a scholarship that covered my travel expenses to participate at “28th International Workshop on Retroviral Pathogenesis” held in 2016 in New Orleans, Louisiana, USA.

This work was supported by

- NÁVRAT (LK11215) - MŠMT ČR
- GAUK 173215
- institutionally by RVO: 68378050

I would also like to acknowledge people and institutions kindly providing the material:

- Primary red deer and fallow deer cells by Ramsés Reina Arias.
- FeLV vectors used for controls by Kazuo Nishigaki.
- The HEK293T cells producing infectious virus with Amphotropic envelope by Stanislav Indik.
- PERV- related material by Magda Matoušková.
- Bat samples by Pavel Hulva and Petr Benda.
- Red deer samples by Institute of Animal Science, Praha - Uhříneves.
- Mule deer NGS data by Mary Poss.

Finally yet importantly, I would like to thank my husband Miroslav for an enormous amount of love and support he provides and the rest of my family for supporting me as well.

ABBREVIATIONS

ADAR- adenosine deaminase acting on RNA
AID- Activation-induced cytidine deaminase (also known as AICDA)
APOBEC- apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
ASLV- Avian sarcoma leukosis virus
BLAST- Basic local alignment search tool
BLASTp- BLAST for proteins
BLV- Bovine leukemia virus
C10- HEK293T cells producing CrERV and GFP
CA- capsid
CAEV- Caprine arthritis encephalitis virus
cGAMP- cyclic guanosine monophosphate–adenosine monophosphate
cGAS- cyclic GMP-AMP synthase
CHOK1- Chinese hamster ovary cells
CHRNA1- cholinergic receptor nicotinic alpha 1 subunit
Cr5- HEK293T cells stably transfected with molecular clone of CrERV
CrERV- cervid endogenous retrovirus
CrERV-IND- CrERV induced by cocultivation
CrERVmut- CrERV with mutation in pro
CRISPR-Cas- Clustered regularly interspaced short palindromic repeats-CRISPR associated sequences
CVO- Cynocephalus volans ELVgv
CypA- cyclophilinA
cytB- cytochrome B
DAPI- 4',6-diamidino-2-phenylindole
DKV- deer kidney virus
DMEM - Dulbecco's modified Eagle's medium
DMSO- dimethyl sulphoxide
dN- number of non-synonymous mutations
DnERV- endogenous Betaretrovirus in *Dasypus novemictus*
dS- number of synonymous mutations
EDTA- ethylenediaminetetraacetic acid
EIAV- Equine infectious anemia virus
ELFO- electrophoresis
ELVgv- endogenous lentivirus of *Galeopterus variegatus*
ELVmpf- endogenous lentivirus of Mustela putorius furo ELVmpf
enJSRV- endogenous Jaagsiekte sheep retrovirus retroviruses
ENTV- Enzootic nasal tumor virus
env- envelope glycoprotein
envvT- truncated envelope glycoprotein
ERV- Endogenous retrovirus
ERV-DC- endogenous retrovirus- domestic cat
ERVW1- endogenous retrovirus group W member 1
FACS- Fluorescence-activated cell sorting
FeLIX- truncated envelope protein encoded by an endogenous FeLV
FeLV- Feline endogenous leukemia virus
FeLV-D- Feline endogenous leukemia virus domestic
FES- FES proto-oncogene, tyrosine kinase
FIV- Feline immunodeficiency virus
Fv- Friend virus susceptibility
Gag- group antigen,
GaLV- Gibbon ape leukemia virus
GFP- green fluorescent protein
GVA3- Galeopterus variegatus specimen 3
GVA5- Galeopterus variegatus specimen 5
GVAgb- Galeopterus variegatus, GenBank sequence
HBZ- HTLV bZIP factor
HEK293T- Human embryonal kidney 293T cells
HERV- Human endogenous retrovirus
HIV- human immunodeficiency virus
HTLV- human T cell leukemia virus
IFNα- interferon alpha
IN- integrase
JSRV- Jaagsiekte sheep retrovirus
KoRV- Koala retrovirus
LTR-long terminal repeat
MA- matrix,
MAFFT- multiple alignment using fast fourier transform
MAVS- mitochondrial antiviral signaling protein
MbRV- Melomys burtoni retrovirus
MINERVa Miniopterus endogenous retrovirus
ML- maximum likelihood
MLV- Murine leukemia virus
MMTV- Murine mammary tumour virus
Mtv– endogenous MMTV
MuLV- Murine Leukemia Virus
MUSCLE- multiple sequence comparison by log-expectation
MW- molecular weight
Mx- Myxovirus resistance
MYA- million years ago
NC- nucleocapsid;
NCBI- National Center for Biotechnology Information
Nef- negative factor
NFκB- nuclear factor κB
NGS- next generation sequencing
NJ- neighbor joining
OERV- ovine endogenous retroviruses
OHK- Odocoileus hemionus kidney
ORF- open reading frame
PBS- phosphate buffered saline
PBS- primer binding site
PCR- polymerase chain reaction

PERT- product enhanced RT assay
PERV- porcine endogenous retrovirus
PERV- porcine endogenous retrovirus
PFM- paraformaldehyde
Pika-BERV- pika-beta endogenous retrovirus
Pol- polymerase
PPT- polypurine tract
PR- protease,
pSIV- endogenous simian immunodeficiency virus
pTEFb- positive transcription elongation factor
PtERV- Pan troglotydes endogenous retroviruses
PyERV- python morulus endogenous retrovirus
qPCR- quantitative PCR
RAG2- recombination activating gene 2
Refrex- soluble restriction factor against feline endogenous and exogenous retroviruses
REL- random effects likelihood
RELIK- Rabbit endogenous Lentivirus type K
RELIK rabit endogenous lentivirus type K
Rev- regulation of expression of virion proteins
Rex- protein encoded by deltaretroviruses, escorts unspliced and singly spliced RNAs out of the nucleus of infected cell
Rmcf- Resistance to mink cell focus-forming virus
RPM- rotation per minute
RSV- Rous sarcoma virus
RT- reverse transcriptase,
SA- splice site acceptor
SAMHD1 Sterile alpha motif and histidine-aspartate domain containing protein 1
SD- splice site donor
SDS- sodium dodecyl sulphate
SDS-PAGE- sodium dodecyl sulphate-polyacrylamid gel electrophoresis
SIV- simian immunodeficiency virus
SloEFV- Sloth endogenous foamy virus
SRA- sequence read archive
STING- stimulator of interferon genes
STLV- simian T-cell leukemia virus
SU- surface subunit of env
TAR- trans-activator response element
Tat- trans-activating regulatory protein
Tax- transcriptional activator that activates viral and cellular genes
tBLASTn- translated BLAST nucleotide
TM- transmembrane subunit of env
TrEMBL- Translated European Molecular Biology Laboratory
TREX1- three prime repair exonuclease 1
TRIM5α- tripartite motif alpha
TSD- target site duplication
UrsusERV- Ursus Endogenous Retrovirus
Vif- virus infectivity factor
Vpr- Viral protein regulatory
Vpu- Viral protein unknown
Vpx- Viral protein x
WDSV- Walleye dermal sarcoma virus
WEHV- Walleye epidermal hyperplasia virus
WGA- whole genome amplification
WGS- whole genome shotgun
Xen1- Xenopus laevis endogenous retrovirus

ABSTRACT

Endogenous retroviruses (ERVs) originate by germline infection and subsequent mendelian inheritance of their exogenous counterparts. With notable exceptions, all mammalian ERVs are evolutionarily old and fixed in the population of its host species.

Some groups of retroviruses were believed not to be able to form endogenous copies. We discovered an additional endogenous Lentivirus and a first endogenous Deltaretrovirus. Both of these groups were previously considered unable to form endogenous copies. Endogenous lentiviruses were discovered only recently and are still quite rare. These are still just small pieces of evidence insufficient to give a broader picture about the history of virus endogenization. We described a novel endogenous Lentivirus in the genome of Malayan colugo (*Galeopterus variegatus*) denoted ELVgv (endogenous Lentivirus of *G. variegatus*). Based on several analyzes we proved that this is the oldest Lentivirus discovered up to date and confirmed its presence in the only other extant species of Dermoptera - *Cynocephalus volans*.

Endogenous deltaretroviruses were the last group without a single endogenous member. We detected the remnants of endogenous Deltaretrovirus in the genome of Natal Long-fingered bat (*Miniopterus natalensis*). However, this sequence was present in the genome only in one copy. We subsequently amplified and sequenced the provirus remnants from other related Miniopteridae bats.

Besides filling in the gaps of missing types of endogenous retroviral copies in genomes, we tried to add to current knowledge about the process of endogenization. The processes accompanying endogenization and the features of viruses capable of endogenization are still not well elucidated.

This might be owed to absence of a suitable model of endogenization. We propose such a model. Besides endogenous retrovirus in koalas, ERV in mule deer (*Odocoileus hemionus*) forms new germ line insertions in the natural host population in the present evolutionary time and might serve as an important model of the retrovirus endogenization process. We have determined complete genome sequence of the deer ERV, denoted cervid endogenous retrovirus (CrERV). In the previous studies, thousands of highly polymorphic CrERV integrations in approximately 50 animals were

characterized. Notable polymorphism within the population of mule deer with CrERV integration sites allocated to specific area verify the predicted young age of the virus as well as the current process of endogenization.

We performed experiments to characterize CrERV from virological perspective and explain the inefficiencies in virus replication cycle, for CrERV exhibits xenotropic behavior despite being efficient in creating new germ line copies. Experiments tackling this question were only partially successful and several questions remained unanswered. Besides these experiments, we tried to assemble retrovirus restriction factors from Cervidae species' genomes and perform analyzes to estimate possible presence of their positive selection.

We also came across of concept, which could elucidate the occurrence of a replication block of viruses with amphotropic envelope in Chinese hamster ovary cells (CHOK1). We propose that these cells (widely used in biotechnology applications) bear an endogenous retrovirus unable to produce infectious particles, but able to produce defective Env protein. This protein might inhibit infection by exogenous retrovirus by competitive inhibition at the receptor.

ABSTRAKT

Endogenní retroviry (ERV) vznikají retrovirovou infekcí zárodečné linie a následným přenosem do dalších generací podle pravidel Mendelovy dědičnosti. Až na pár výjimek jsou všechny druhy savčích ERV evolučně staré a fixované v populaci svých hostitelských druhů.

O některých skupinách retrovirů se předpokládálo, že nejsou schopny vytvářet endogenní kopie. Objevili jsme další příklad endogenního Lentiviru a první endogenní Deltaretrovirus. Obě tyto skupiny byly dříve považovány za neschopny vytvářet endogenní kopie. Endogenní lentiviry byly objeveny pouze nedávno a stale se považují za velmi vzácné. Toto jsou stále jen minoritní důkazy z kterých nemůžeme získat celkový obraz o průběhu virové endogenizace. Popsali jsme nový endogenní Lentivirus v genomu letuchy malajské (*Galeopterus variegatus*) a nazvali ho ELVgv (endogenous Lentivirus of *G. variegatus*). Na základě několika analýz jsme dokázali, že se jedná o nejstarší dosud objevený Lentivirus, a potvrdili jsme jeho přítomnost v jediném jiném současném druhu Dermopter - *Cynocephalus volans*.

Endogenní deltaretroviry byly poslední retrovirovou skupinou bez nalezeného endogenního člena. Našli jsme zbytky endogenního Deltaretroviru v genomu netopýra létavce natalského (*Miniopterus natalensis*). Tato retrovirová sekvence byla přítomna v genomu pouze v jedné kopii. Následně jsme tento provirus amplifikovali pomocí PCR a osekvenovali také z jiných příbuzných druhů čeledi Miniopteridae.

Kromě vyplnění mezery v typech endogenních retrovirových kopií nalezených v hostitelských genomech jsme se dále snažili rozšířit současné poznatky o vlatním procesu retrovirové endogenizace. Procesy doprovázející endogenizaci a vlastnosti virů schopných endogenizace nejsou stále dostatečně objasněny.

To je částečně zapříčiněno chybějícím vhodným modelovým systémem pro endogenizaci. Jako vhodný model navrhujeme ERV jelence ušatého (*Odocoileus hemionus*), který vytváří nové inzerce v zárodečných buňkách v populaci svého hostitele v nedávne evoluci, a může sloužit jako důležitý model pro studium procesu retrovirové endogenizace. Popsali jsme kompletní sekvenci genomu jelenčího ERV, nazvaného cervid endogenous retrovirus (CrERV). V předchozích studiích byly charakterizovány

tisíce vysoce polymorfních integrací CrERV v přibližne 50 zvířatech. Tento pozoruhodný inzerční polymorfismus v populaci severoamerických jelenců, s jednotlivými integracemi CrERV typickými pro konkrétní geografické lokality, naznačuje předpokládaný mladý evoluční věk viru a také současně probíhající proces endogenizace.

Provedli jsme experimenty zaměřené na charakterizaci CrERV z virologického hlediska a na vysvětlení bloků retrovirového replikačního cyklu, protože CrERV vykazuje xenotropismus navzdory efektnímu vytváření nových zárodečných kopií. Tyto experimenty byly úspěšné jemom zčásti a mnohé otázky zůstaly stále otevřené. Kromě těchto experimentů jsme se dále snažili sestavit sekvence retrovirových restrikčních faktorů z genomů jelenovitých druhů a provést analýzy na možnou přítomnost pozitivního selekčního tlaku

Dále jsme se pokusili objasnit přítomnost replikačního bloku virů s amfotropní retrovirovou obálkou na linii ovarialních buněk čínského křečka (Chinese hamster ovary cells - CHOK1). Předpokládame, že tyto buňky (hojně využívané v biotechnologii) v sobě nesou endogenní retrovirus neschopný produkce infekčních částic, ale schopný produkce a sekrece defektního obalového glykoproteinu (Env). Tento glykoprotein může působit inhibičně na infekci exogenním retrovirem, mechanismem kompetitivní inhibice na virovém receptoru.

CONTENTS

1 HYPOTHESES AND AIMS

1) The first aim of this thesis is to characterize the findings of the computational screening of all publicly available genomes. This screen was aimed at the discovery of novel or unusual endogenous retroviruses. We chose two hits from this screen to be characterized in this work.
    - Discovery of an endogenous Lentivirus in *Galeopterus variegatus* (Malayan colugo) denoted ELVgv (Endogenous Lentivirus in *Galeopterus variegatus*).
        - Characterize the orthologous and paralogous sequences found in *Galeopterus variegatus* and the only other extant species from Dermoptera - *Cynocephalus volans.*
        - Characterize the relationship of ELVgv and its host.
    - Discovery of the first endogenous Deltaretrovirus found in the genomes of Miniopteridae bats denoted MINERVa (Miniopterus endogenous retrovirus).
2) The second aim of this thesis was to induce Cervid endogenous retrovirus (CrERV) from mule deer cells by cocultivation with susceptible human cells to characterize virus by virological methods.
3) The third aim was to describe the host-virus interactions of CrERV. The original idea was to use gammaretroviral pseudotypes to identify the replication block of the virus in mule deer cells. This approached was later complemented with marker rescue assay.
4) The fourth aim is the further description of host-virus interactions of CrERV by assembling the host retrovirus restriction factors *in silico* and estimating the magnitude of the positive selection towards them.
5) The fifth aim of the study was to analyze whether the infection block of Chinese hamster ovary (CHOK1) cells occurs due to the presence of endogenous retrovirus fragments secreted by CHOK1 cells.

Figure 1 schematically shows the planned pseudotype constructs (MLV core and CrERV envelope and vice versa) and what would the infectivity outcomes on mule deer or human cells indicate.

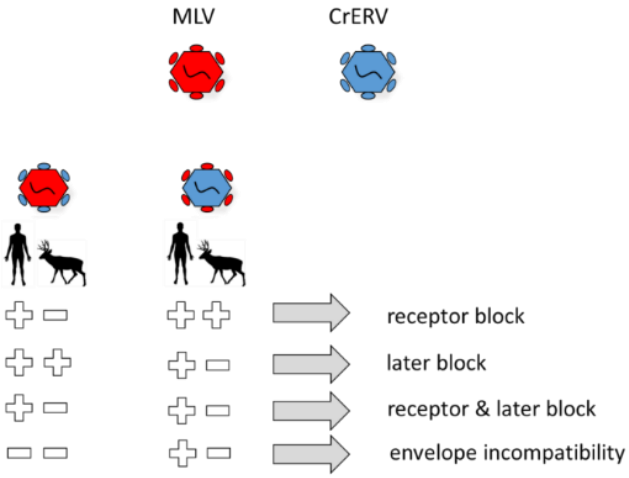

Figure 1: Possible outcomes of the pseudotyping experiment. The parts used from MLV virus are depicted in red and parts of CrERV virus are depicted in blue. Hexagon indicates gag-pol and ovals indicate env. The infectivity is depicted by + (susceptible) or – (resistant).

2 INTRODUCTION

2.1 Retroviruses

Retroviruses are the causative agents of various pathologies (e.g. tumors, immunodeficiency, and neurological disorders). Retroviruses compose roughly 10% of mammalian genomes, composing approximately 50% of genomes together with the other retroelements. Retroviruses are enveloped single-stranded RNA viruses containing reverse transcriptase enzyme. Reverse transcriptase enables retroviruses to be an exception in the central dogma of molecular biology, being a key player in transcribing their RNA genome to DNA (hence the name retro-viruses). They are used as a molecular biology tool and their close examination brought several important discoveries (carcinogenesis, cellular growth control, oncogenes, signal cascades and various other issues in molecular biology). Therefore, the continuous interest in them and close examination of their possible exploitation is still an issue. This is also exemplified by the fact that retroviral vectors have been used in more than 300 clinical trials.

2.1.1 Retroviral structure

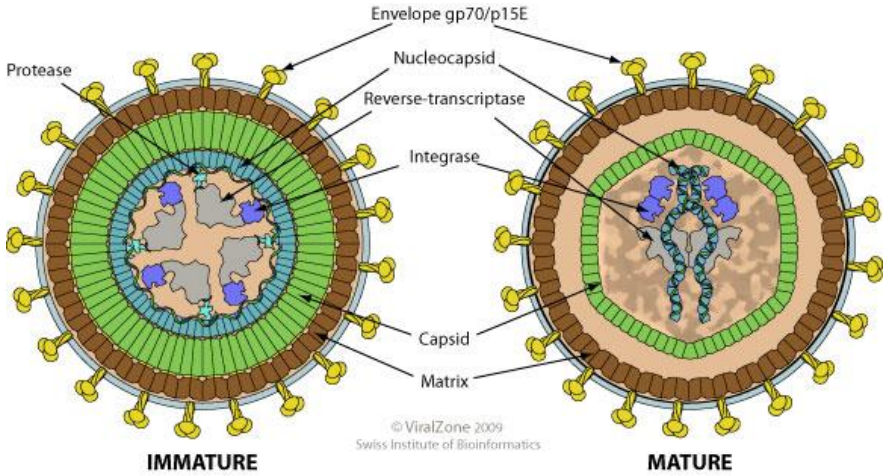

Figure 2: Schematic structure of a retrovirus (Gammaretrovirus). The pictures depict schematic structure of an immature and mature virion. The Env is depicted as a trimer of a surface (SU) and transmembrane (TM) subunit. Pictures are adapted from webpages of Swiss institute of Bioinformatics (http://viralzone.expasy.org/).

The three major genes in the retroviral genomes are *gag* (group antigen), *pol* (polymerase) and *env* (envelope glycoprotein). Proteins coded by *gag* are MA (matrix), CA (capsid), NC (nucleocapsid); proteins coded by *pol* are PR (protease), RT (reverse

transcriptase), IN (integrase); and the *env* gene codes for the SU (surface) and TM (transmembrane) subunits of the retroviral envelope. Retroviruses have a pseudodiploid RNA genome encapsulated in the viral core. Schematic structure of a retrovirus is depicted in Figure 2. The structure of the cores and genomes varies among retroviral genera.

2.1.2 Retroviral life cycle

Retroviral life cycle consists of early phase (from entry into the cell up to integration into the host genome) and late phase (expression of the integrated retrovirus, assembly, and release of the retroviral particles). During several steps of this process retroviruses hijack the host molecular machinery for their replication. Starting from entry, which is enabled via host surface proteins (virus receptors), then continuing with exploitation of host transcription and translation complexes to acquiring lipid bilayer from host cells.

Entry into the cell is the first step in the retroviral life cycle. The viral Env interacts with the cellular receptor and this induces conformational changes in the transmembrane Env subunit (TM). Many examples demonstrate that the interaction of cellular receptor and retroviral Env might not be sufficient for the retroviral entry into the cells (e.g. presence of co-receptors is needed). However, cellular receptor and Env interaction is still believed to be the most crucial aspect of the cellular entry.

Reverse transcription occurs right after the viral particle enters the cytoplasm. The process of reverse transcription was discovered while studying Rous sarcoma virus (RSV) (reviewed in (Baltimore, 1995). It is dependent on the two activities of reverse transcriptase: DNA polymerase (able to utilize both, DNA and RNA as template) and a nuclease (ribonuclease H). The product of the process is a double stranded DNA genome.

Reverse transcription starts with one RNA of the pseudodiploid genome, utilizing as a primer a cellular tRNA bound to PBS (primer binding site). The elongation continues towards the 3‘end (left RNA LTR-long terminal repeat) synthetizing the left DNA LTR. When the RNA template for the elongation ends, the original RNA LTR is digested and the newly synthesized DNA part of the LTR binds to the right (downstream) RNA LTR.

Subsequently elongation continues until the newly synthesized first DNA strand reaches the PBS used for the first binding of the primer. The remaining RNA is digested except for PPT (polypurine tract) located right next to the downstream LTR. The remaining RNA serves as a primer for the synthesis of the second DNA strand. When the synthesis reaches right LTR, the DNA double strand disassociates and the newest shorter strand (already without PPT) serves as a primer for a full length double strand viral DNA synthesis. This results in the fact that after the process of reverse transcription is over, the dsDNA has a complete LTR (consisting of U3-R-U5 sequences) on both sides. Retroviruses can recombine during reverse transcription by template switching during DNA synthesis. After mixed infection, half of the produced retroviruses are recombinants (Goodrich & Duesberg, 1990).

Some retroviruses are known to be able to actively enter the nucleus. They are heavily disassembled in the host cell cytoplasm and are subsequently able to exploit the host nuclear transport machinery (e.g. HIV). Other retroviruses have to wait for the division of the cell nucleus in the cell cycle when the nuclear envelope is temporarily disassembled (e.g. MLV). These processes were reviewed recently (Cohen, Au, & Pante, 2011).

Prior to integration, a pre-integration complex is formed. It is a complex of host and viral proteins which can be isolated from infected cells (Farnet & Haseltine, 1990). This complex enables performance of three steps of the integration: processing, joining, and repair. Processing of the retroviral DNA ends (cleavage of two nucleotides from both ends) and joining of viral and host DNA is mediated by viral protein integrase. The integration is finalized in the last step when occurring gaps in DNA strands are filled by host polymerases. The result of the retroviral life cycles up to this step is integrated viral DNA (called provirus) with identical LTRs on both ends ready for transcription. Upon integration, short target site duplication (TSD) of host DNA is formed.

Transcription produces RNA templates for subsequent translations of retroviral genes as well as full RNA genomes to be later packed into the newly released viral particles. Transcription is performed by the host transcriptional machinery (RNA polymerase II). The organization of the genomes of particular retroviral genera with weak and strong stop codons enables translation of particular genes to occur in desired

orders and amounts. For example, lentiviruses have a weak stop codon ending the *pol* sequence which enables the read-through transcription of *gag* or *gag-pol*. The *env* is transcribed after a read-through and a frameshift, for Env is needed in lesser amount for retroviral assembly. Splicing occurs abundantly and is present also in simple retroviruses, generally for *env* (see Figure 2 for reference), and in complex retroviruses for additional accessory genes. The following translation must occur in the right moment in order to produce sufficient amount of particular gene product, the last being the products of *gag* and *env* to prepare enough protein for the retroviral particle assembly.

The process of retrovirus assembly occurs in the cellular cytoplasm close to the cytoplasmic membrane. The site of assembly varies for different retroviruses. When the retrovirus assembly is complete, the virus buds from the cell and finally is released. These processes are determined mainly by the products of the *gag* gene. Encapsidation signals on the unspliced RNAs are recognized by nucleocapsid proteins and the encapsidated RNA is then not to serve as a transcription template.

When the retrovirus is released from the cell, the particles undergo so-called maturation. This step enables the particles to become infectious for further processing of Gag and Gag-pro-pol is required. Mutants in the retroviral PR (protease) domain or particles assembled in the presence of PR inhibitor are not able to undergo this step. For example, gammaretroviral R-peptide (prevents fusion with host cell membranes before budding) needs to be cleaved in order to create efficient Env protein (Schneider et al., 2011).

2.1.3 Retrovirus classification

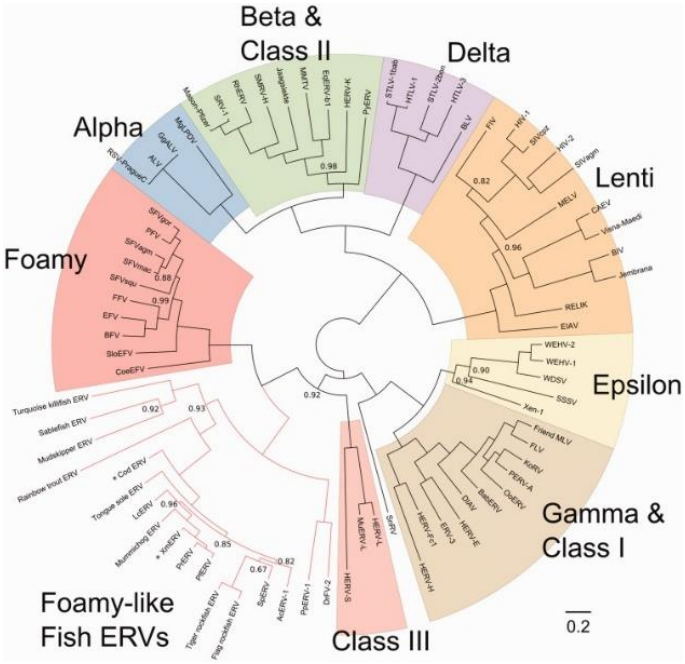

Figure 3: Phylogenetic relationship of retroviral groups (adapted from (Ruboyianes & Worobey, 2016).

The retroviral family is composed of seven genera (see Figure 3). *Alpharetroviruses*, *betaretroviruses*, and *gammaretroviruses* are simple retroviruses encoding only basic retroviral genes and mostly lack accessory genes. *Deltaretroviruses*, *epsilonretroviruses*, and *lentiviruses* are complex retroviruses coding for several accessory genes. *Spumaviruses* (Foamy viruses) are a special clade of retroviruses that generally do not cause pathologies.

2.2 Genomics state of art as a defining factor in discovering endogenous retroviruses

The current state of the overall sequenced genomes is the major factor in determining the advances in the endogenous retroviruses research. The number of currently sequenced animal genomes available at NCBI (National Center for Biotechnology Information) is 4 amphibians, 73 birds, 80 fishes, 32 flatworms, 82 roundworms, 222 insects, 127 mammals, 16 reptiles, and 83 other unclassified genomes (data from January 2017). Importantly, the assemblies of these genomes are of a

variable quality. ERVs, which are mostly present in many highly similar copies, are often either missing or assembled and annotated incorrectly in the genomes.

The cost of sequencing has been decreasing rapidly during the past few years and this might have been a factor in new initiatives such as Genome 10K project. This project aims to sequence ten thousand vertebrate genomes (Genome, 2009). One of the first groups of sequenced animals is 48 avian species (SJ, Haussler, & Ryder, 2014). In two years since the project was launched, the number of sequenced genomes increased from 26 to 277 (Koepfli, Paten, Genome, & O'Brien, 2015). The project was enlarged and gave rise to an initiative by Avian phylogenomics consortium to sequence ten thousand bird genomes (Birds10K) by the year 2020 (Zhang et al., 2015). Other large-scale projects such as Bat1K (Skibba, 2016), which aims to analyze bat communication via sequencing thousand bat genomes, provide set of sequencing data which can be further utilized to screen for endogenous retroviruses.

The rapid accumulation and development of Next generation sequencing (NGS) data sets available for public use from various species enables everyone with required skills to screen for sequences of their interest. Besides all possibilities, it potentially provides data to deepen our understanding of evolution of retroviruses. Evolutionary events in extant and recently infecting retroviruses combined with population genetics might elucidate the biology of retroviruses from various aspects, including endogenization (Johnson, 2015). Other areas of biology could benefit from sequencing and assembling various animals' genomes as well. The technology designed for sequencing human genome and other large scale projects (such as 10,000 human genomes (Genomes Project et al., 2010) and ENCODE (http://www.encodeproject.org)) might be further utilized in this aspect to fill in the gaps of understanding major biology questions in various fields (Richards, 2015).

2.3 Retrovirus endogenization

Endogenous virus is a virus which infected and integrated into the germline cells and is further inherited vertically in mendelian fashion. The process of establishing the presence of a virus in the host cell is called endogenization. The term 'endogenous retrovirus' can be used to denote both, the integrated DNA sequence, and the infectious

particles produced by this sequence. The fate of the endogenous retrovirus in the infected population might vary. Generally, three possible outcomes are present. First, the virus might become fixed in the population; second, it might remain in a polymorphic state; and third, it might get washed out from the population completely (Katzourakis, Rambaut, & Pybus, 2005).

The presence of the endogenous retrovirus might provide an advantage to the infected host. This might result in an enhanced spreading of the endogenous retrovirus element and a higher probability of its fixation in the host population. This scenario is probably rare, with best example provided by the presence of *Syncytin* gene originating from a retrovirus infection among placental animals. *Syncytin* produces fusogenic protein utilized in placentation (discussed in the chapter 2.6.3 Endogenous retroviruses with a role in the host physiology).

The presence of the endogenous retrovirus might not have an impact on the host whatsoever. Such neutrality is presumably the most common case. These proviruses might as well get fixated or vanish from the population, as well as they might be present in the genome in the polymorphic integration in the host genomes. While there is neither positive nor negative selection towards the fixation of an endogenous retrovirus in the host genome is rather random (Rouzine, Rodrigo, & Coffin, 2001). Such a situation is desired in the model used to study processes accompanying endogenization on a population scale.

The presence of the endogenous retrovirus might be a burden for the host. The product of the virus might be toxic or even lethal for the host. The major disadvantage of the presence of the provirus for the host occurs if the provirus integrates into a gene exon or into intron in the plus orientation and therefore disrupts it or causes aberrant splicing. Such a provirus will be usually lost from the population early after integration (van de Lagemaat, Medstrand, & Mager, 2006).

All of the three possible scenarios of retroviral spread in the population are depicted in the Figure 4.

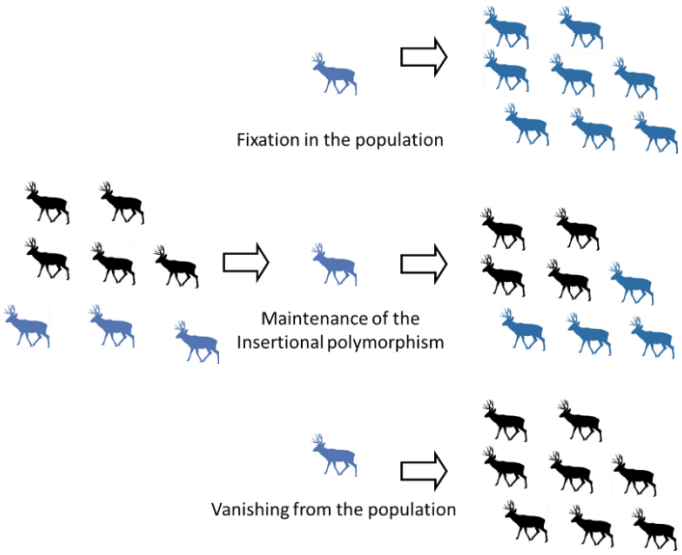

Figure 4: The possible fate of endogenous retrovirus integration in the population. The virus might have no remarkable impact on the host, so the spread of the virus in the population is affected by other non-evolution related events (e.g. geographic barrierrs, bottleneck); leading to different integration patterns in different sub-populations.

Despite the fact that retroviruses were believed to be the only viruses capable of endogenization, further examination of sequences suggested that this process is not exclusive to them (see Table 1 for reference). This illustrates that retroviruses might not be the only viral elements shaping the animal genomes.

Table 1: List of remnants of non-retroviral viruses among animal genomes.

Positive sense single strand RNA viruses		
Flaviviruses	Yellow fever mosquito	(Katzourakis & Gifford, 2010)
Negative sense single strand RNA viruses		
Bornaviridae	Ground squirrel Snakes Thirteen-lined ground squirrels Bats	(Horie et al., 2010) (Gilbert et al., 2014) (Suzuki, Kobayashi, Horie, & Tomonaga, 2014) (Cui & Wang, 2015)
Double strand RNA viruses		
-	-	-
Viruses utilizing reverse transcriptase in their life cycle		
Hepadnaviruses	Passerine birds	(Katzourakis & Gifford, 2010)
Single strand DNA viruses		
Dependoviruses	Domestic dog, Guinea pig, Nine-banded armadillo, Horse, Tammar wallaby, African elephant, Mouse, Little brown bat, Pika, Duckbilled platypus, European rabbit, Hamadyras baboon, Cape hyrax, Malayan flying fox, Brown rat, Bottlenose dolphin, Alpaca	(Feschotte & Gilbert, 2012)
Parvoviruses	Guinea pig, Tenrec, Rat, Tammar wallaby, Opossum	
Amdoviruses	Cape hyrax	
Circoviruses	Domestic dog, Cat, Giant panda, and Opossum	
Double strand DNA viruses		
Herpesviruses	Human, Aye-aye, Bonobo, Philippine tarsier	(Aswad & Katzourakis, 2014)
Pappillomaviruses	Platypus	(Cui & Holmes, 2012b)

2.3.1 Estimating the age of integrated ERV elements

Endogenized viruses are a valuable tool in studying deeper evolutionary history of viruses. Upon endogenization the virus genome starts to mutate at a much slower rate (being the mutation rate of the host mammalian genome) than exogenous retrovirus, this difference can be up to 10^6-fold. Thanks to this fact, we are able to connect events in recent and ancient viral evolution (Aiewsakun & Katzourakis, 2015).

The time of the integration of ERV elements into the host genome might be estimated by several approaches. 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 species bearing the ERV studied. This method is very robust, but yields usually a quite broad time interval for the ERV age estimated.

A second method is dubbed 'LTR aging'. This approach exploits the fact that both LTRs are identical at the time of integration, and uses the number of sequence differences that occurred since that time until present. The calculation of the time (T) needed to accumulate a given number of sequence differences (N) in the combined LTR length (L), assuming e.g. a neutral genomic substitution rate (R) of 2.3×10^{-9} to 5×10^{-9} per site per year, following formula: $T = N/(R \times L)$ can be used for the age estimation of the ERV integration (Johnson & Coffin, 1999). This provides additional data about dating individual viral groups as well as it might serve as an additional marker in dating host species divergence. However, limitations of this method such as presence of a single LTR, recombination, gene conversion, and probable differences in the mutation rate at different sites must be taken into account.

A third method uses the fact that genomic loci with ERV integration might duplicate in the evolutionary history of the host species. If the virus is present in both regions, then the duplication event must have occurred after virus integration and a minimal time estimate can be obtained (Hron, Fabryova, Paces, & Elleder, 2014; Katzourakis, Tristem, Pybus, & Gifford, 2007).

The fourth approach is the time-calibrated phylogenetic analysis of orthologous proviral sequences from multiple species (using molecular clock). This is the most sophisticated method and provides information about the whole ERV lineage, not just about specific ERV integrations (Jha et al., 2009; Kamath et al., 2014; Tonjes & Niebert, 2003).

All of the methods used in estimating the age of integration of the endogenous retrovirus have limitations. Statistically speaking, the used method to estimate the time of the integration must 'fit the data' (Shapiro et al., 2011). For example, analyzes of extremely old retroviruses (pushing their origin to more 450 MYA) must often consider the accuracy of the methods when it comes to such old retroviruses (Aiewsakun & Katzourakis, 2017).

2.3.2 Proposed models for studying endogenization of retroviruses

The process of endogenization and the events accompanying it are still poorly elucidated. As reviewed previously, endogenization might occur among virus families other than retroviruses, but rather rarely. Endogenous retroviruses (with emphasis on the rare ones) and some of their exogenous analogues are described in the next chapter. Here, the viruses perceived to have a potential to elucidate the events accompanying endogenization are described.

2.3.2.1 Koala retrovirus (KoRV)

One of the possible models for studying the process of endogenization is a recently discovered virus in Australian koalas, Koala endogenous retrovirus (KoRV). Its integration polymorphism (the presence of a proviral DNA at a particular integration site only in some individuals and a complete presumed lack in others) is a reason to consider it a young endogenous retrovirus (Tarlinton, Meers, & Young, 2006). Despite the fact, that insertional polymorphism was described already in other species (e.g. mice (Frankel, Stoye, Taylor, & Coffin, 1990), cats (Banerji, Kapur, & Kanjilal, 2007), sheep (Chessa et al., 2009)), so far it was always only a small number of integrations. The process of endogenization and subsequent adaptation in the host genome is still not elucidated.

KoRV particles were first described in the tissues of leukemic koalas (Canfield, Sabine, & Love, 1988). The virus was sequenced and detected in stimulated peripheral blood cells as well as in three leukemia-positive koalas. The virus displayed sequence similarity (78% sequence identity) to Gibbon ape leukemia virus (GaLV) (Hanger, Bromham, McKee, O'Brien, & Robinson, 2000). The elevated level of KoRV transcripts in leukemic koalas is suggestive to the fact that KoRV causes neoplastic diseases in koalas (Tarlinton, Meers, Hanger, & Young, 2005). However, no causative studies have been performed yet.

The sequence similarity of GaLV and KoRV remained an issue to resolve. An endogenous virus in Melomys burtoni might be the link between these two viruses (Simmons, Clarke, McKee, Young, & Meers, 2014). An insight into this problem was given by comparison of GaLV, KoRV, and Murine leukemia virus (MLV with Amphotropic

envelope 4070A) infectivity and receptors. These studies indicate that the host range of GaLV and KoRV do overlap, but their envelopes alter. Both GaLV and KoRV are capable of infecting a wide species range (Oliveira, Farrell, & Eiden, 2006).

There was another possible mode of transmission of KoRV described. The possible vertical transmission of the retrovirus combined with the fact that KoRV copy number varies among individuals suggests that the virus is probably invading the genome of koalas. The prevalence of KoRV varies based on the geographic region. None of the samples obtained from Kangaroo Island (Australia) were positive. This enabled us to witness the initial entry of the endogenizing retrovirus into the wildlife species population (Tarlinton et al., 2006).

The phenomenon of varying KoRV presence was also observed when koalas kept in ZOOs were analyzed (reviewed in (Denner & Young, 2013)). This fact combined with the presence of currently active exogenous variants of KoRV suggested to cause pathologies in koalas makes it an interesting platform to study the process of endogenization. The significance of the virus is emphasized by the veterinary importance (reviewed in (Kinney & Pye, 2016)).

However, the usage of KoRV as a model to study endogenization has its pitfalls. Attempt to characterize its integration sites occurred only recently. The results indicate that the koala genome might have been invaded by KoRV at least seven times with the most recent integrations up to 50,000 years ago (Ishida, Zhao, Greenwood, & Roca, 2015) The samples obtained from living koalas were compared to the historic museum samples (Avila-Arcos et al., 2013). The results indicate that only a small number of KoRV integrations sites recognized as recently integrated are shared by multiple animals. The regional differences in KoRV fixation were proposed, despite the fact that only small number of animals was analyzed. Beside that, the genome of koala is still not assembled and released, therefore the closest genome (Tammar wallaby - *Macropus eugenii*) is often used for KoRV analyzes (Cui et al., 2016).

2.3.2.2 Cervid endogenous retrovirus (CrERV)

A situation similar as the one occurring in koala genome is also occurring in the genome of North American mule deer (Kamath et al., 2014). Despite the virus being

already described (Aaronson, Tronick, & Stephenson, 1976), only current methods enabled its thorough examination.

Initially, CrERV was described as the first endogenous retrovirus in the a species originating in the New World. It was described to possess distinct immunological properties. The endogenous retrovirus was induced by cocultivation with both, human and equine cells. The virus was described to exhibit xenotropic behavior. In most of the studies connecter to CrERV, the struggle with low virus titers was present (Aaronson et al., 1976).

At the time of the initial description of the virus, sequencing techniques were not widely avaliable. The tools available to use to determine the endogenous retrovirus sequence similarity to other endogenous retroviruses of other members of related phyla (such as Artiodactyla) were mainly hybridization techniques. The obtained CrERV sequence at the time hybridized strongly with the members of Cervidae (approximately 85% similarity) clade, less effectively with the Bovidae clade (approximately 20% similarity) and did not hybridize with the related virus sequences obtained from more distant animal clades (Tronick, Golub, Stephenson, & Aaronson, 1977).

After more than 30 years, the CrERV caught the eye of researchers again. Its partial sequence was identified in a metagenomics screening of pathogens present in mule deer lymph nodes (Wittekindt et al., 2010). Eventually complete CrERV sequence was obtained and the endogenous nature of the virus was described (Elleder et al., 2012). The presence of specific CrERV integrations in the mule deer and absence in the white-tailed deer suggested that the virus is rather evolutionary young for these two species split approximately 1 million years ago (MYA). Moreover, the provirus was proven to be transcriptionally active (Wittekindt et al., 2010). Using the adapted method based on PCR and next generation sequencing (NGS), hundreds of CrERV integrations were described in the genome of each and every mule deer individual examined (Le Bao, 2014). The integrations were discovered to be highly polymorphic. This means that the germinal cells of mule deer must have been infiltrated by CrERV several times. All this evidence points to the fact that the virus is probably currently endogenizing, thus it seems to be a suitable model to study the processes accompanying endogenization (Elleder et al., 2012).

Further studies of the provirus sequence consisted of sequencing 14 proviruses and following their presence/absence in various mule deer populations. The pattern of the presence of specific proviruses showed that individual CrERV integrations tend to cluster in localized geographic regions. This served as strong independent evidence that these integrations are extremely evolutionarily young, possibly only a few generations of the deer host. The polymorphic nature of CrERV integrations can also be utilized as a powerful genetic marker to study the population structure and history of the host species (Kamath et al., 2014). In general, endogenous retrovirus-derived genetic markers are very powerful, for the following reasons: (I) their ancestral state is known (absence of virus), (II) extremely high number of variants, because retroviruses can effectively target any position in mammalian genome, and in two orientations of integrated provirus, (III) once integrated, the provirus basically cannot be completely excised from the genome (Biek, Drummond, & Poss, 2006).

All of the aforementioned recent studies of CrERV were based on genetic analysis of its sequences, but did not study the virological aspects of the provirus. Replication of the original Aaronson cocultivation experiments was performed (Aaronson et al., 1976). The induced virus (originally named Deer Kidney Virus – DKV – by Aaronson) was shown to be sequentially identical to the sequences of CrERV used in the recent studies. The virus was proven to be xenotropic and its particles sediment in the area of the Iodixanol gradient typical for retroviruses. Replication-competent clone of the virus was constructed and the infection kinetics was described (Fabryova, Hron, Kabickova, Poss, & Elleder, 2015).

2.3.2.3 Jaagsiekte sheep retrovirus (JSRV), Enzootic nasal tumor virus (ENTV), and endogenous retroviruses (enJSRVs)

JSRV is a Betaretrovirus causing infectious lung cancer in sheep flocks. The JSRV, ENTV, and enJSRV have been studied due to their evolutionary interplay and also restriction effect of enJSRV Gag on exogenous JSRV (discussed later). However, presence of endogenous JSRV and currently infecting JSRV is highly suggestive of the fact that JSRV might be currently endogenizing. The oldest copy of enJSRV invaded the sheep genome approximately 7 MYA. The presence of orthologous copies not only in sheep genome, but also in the genome of goats, himalayan thar, and takin indicate, that the

invasion of the genome occured prior to the speciation of the Ovis genus. The presence of another provirus in takin and sheep, but not in goats and thar indicate that another invasion of the genome occured later on in the genus evolution. Positive selection pressure towards enJSRV is indicated, making it tempting to assume, that the presence of enJSRV in the selected ungulate genomes might be an asset for the animals (Arnaud et al., 2007a).

2.3.3 Other evolutionary young endogenous retroviruses

The aforementioned viruses are currently the best characterized models for studying the processes acompanying endogenization. However, they are not the only cases of evolutionary young endogenous retroviruses, and two other examples are presented below.

2.3.3.1 Polar Bear (Ursus maritimus) Endogenous Retrovirus (UrsusERV)

Polar bear genomes harbor retroviruses phylogenetically related to such evolutionary young viruses as PERV (porcine endogenous retrovirus) and KoRV. The provirus detected in the bear genome was overall intact. None of the integrations of the provirus in the genome were orthologous to integrations among bear species and analysis of the UrsusERV LTRs present in the genome indicates, that the bear genome was invaded by a virus forming an endogenous copy at least twice (Tsangaras, Mayer, Alquezar-Planas, & Greenwood, 2015).

2.3.3.2 Unfixed Chimeric Endogenous Betaretrovirus in Armadillo (DnERV)

Endogenous retrovirus with gammaretroviral *env* gene and otherwise Betaretrovirus features was discovered in the genome of armadillo (*Dasypus novemcinctus*). DnERV has not yet reached fixation in the armadillo genome, because insertional polymorphism was detected among Dasypus genus and only haploid copies of it are common (Malicorne et al., 2016).

2.4 Non-human endogenous retroviruses in vertebrates and their exogenous counterparts

Up to date there was no vertebrate genome described to lack retroelements or endogenous retroviruses. Recent massive sequencing of human and animal genomes led to the realization of the fact that almost 10% of mammalian genomes consist of ERVs. For a long time, it was thought that only simple retroviruses (alpha-, beta-, and gammaretroviruses) are able to create endogenous copies. This theory was proven wrong upon the discovery of the first endogenous Lentivirus - RELIK (Katzourakis et al., 2007) and other endogenous complex retroviruses were subsequently described. These chapters summarize some of the endogenous retroviruses in the animal kingdom and their relationships to exogenous counterparts (if existing). With respect to the topics discussed in this thesis, lentiviruses are discussed in more detail.

2.4.1 Alpharetroviruses

Alpharetroviruses as a model system helped to clarify quite a few phenomena of modern molecular biology. Despite playing a crucial role in elucidating many phenomena in the past, interest in them seems to be declining recently. Endogenous copies of retroviruses were first found in avian alpharetroviruses; hence they are included in this chapter.

2.4.1.1 Avian sarcoma leukosis virus (ASLV) and Rous sarcoma virus (RSV)

The ASLV and RSV might be considered one the most important viruses forming the field of retrovirology in twentieth century. The discoveries based on studies of ASLV were awarded the Nobel Prize three times. First, it was discovered, that chicken leukemia can be transmitted from animal to animal by cell-free tissue filtrate (Ellermann and Bang) and cell-free tumor filtrate (Rous, 1910). Peyton Rous was then awarded the Nobel Prize in 1966. Proceeding in the studies of ASLV and RSV it was proven that these viruses contain RNA genomes. The elucidation of the process of reverse transcription and integration of produced viral DNA was discovered by Howard Temin and two years later H. Temin and David Baltimore independently discovered reverse transcriptase. They were both awarded the Nobel Prize for the discovery in 1975.

ASLV and RSV are genetically very similar, but RSV contains a complete ORF for the *src* gene. The discovery of this gene was awarded the Nobel Prize to J. Michael Bishop and Harold E. Varmus for their discovery of "the cellular origin of retroviral oncogenes". Src is a tyrosine kinase, which triggers uncontrolled growth of cells. The virus acquired the *src* gene (denoted *v-src*) from the host cell (cell analogue denoted *c-src*) during its replication. However, cellular and viral *src* genes differ. *v-Src* lacks tyrosine 527 and is therefore constitutively active without possible inhibitory regulation whereas *c-src* is a strictly regulated proto-oncogene active only when required (Czernilofsky et al., 1980; Smart et al., 1981). Similar cases of cellular proto-oncogenes present in other retroviruses are overviewed in the chapter Oncoretroviruses.

ASLV is still currently studied not only for its ability to trigger cancerogenesis, but also for the fact that its subtypes require different receptors for viral entry (see Table 2 for reference). ASLV is also studied to elucidate the restriction processes resulting in avian retroviruses inability to infect mammalian cells (Lounkova et al., 2014). Endogenous copies of retroviruses were first found in avian alpharetroviruses. First, avian leukosis virus in the domestic fowl (Gallus gallus) was discovered, and murine leukemia virus and murine mammary tumor virus in the laboratory mouse (Mus musculus) followed (Weiss, 2006). Transcriptionally active ASLV promoters were found in pathogen free chickens (McNally, Wahlin, & Canto-Soler, 2010).

2.4.2 Betaretroviruses

MMTV and JSRV might not be particularly rare retroviruses, but they are among the suitable models for studying the processes accompanying endogenization. The species present in this genus include JSRV - Jaagsierte sheep retrovirus, MPMV - Mason Pfizer monkey virus, MMTV - Mouse mammary tumour, HML1-10 - Human mouse mammary tumor virus like, Beta like retroviruses, Python-molurus. Selected examples of endogenous betaretroviral species are described in this chapter.

2.4.2.1 Murine mammary tumour virus (MMTV and Mtv)

MMTV is a milk-transmitted retrovirus (Bittner, 1936). Despite the fact that several mice bear an endogenous copy (denoted *Mtv*), and the possible mode of

transmission is also via the maternal milk enables the virus to spread via two routes. The virus transmitted via the exogenous route is capable of creating tumors later in life.

Mtv is a simple retrovirus, but its genome encodes for additional gene s*ag.* Endogenous copies are either able or not able to produce the functional viral particles due to several mutations, but in most cases the reading frame for *sag* remains intact (reviewed in (Holt, Shevach, & Punkosdy, 2013)). The *sag* probably has some beneficial features for the host with its ability to regulate the host immunity response by affecting the nature of the T-cell reservoir (Kang et al., 1994). This results in affecting the mode of other not exclusively viral infections as well. MMTV also encodes for a gene which is a self-regulatory RNA export gene *rem* (Mertz, Simper, Lozano, Payne, & Dudley, 2005).

2.4.2.1.1 Pika-BERV

An endogenous analogue to MMTV was found in the genomes of pikas (*Ochotona* sp.) and denoted Pika-BERV (Pika-beta endogenous retrovirus). The invasion of the genome was dated to be 3-7 MYA and the proviral sequence was described in several Ochotona species (Lemos de Matos et al., 2015).

2.4.2.2 Jaagsiekte sheep retrovirus (JSRV) and its endogenous counterpart (enJSRV)

The sheep genome is colonized with the endogenous copies of JSRV (the copies being denoted enJSRV). This virus was already discussed as a model used for studying the phenomenon of endogenization. Due to the presence of an interesting interplay between JSRV and enJSRV, this virus is discussed in more detail in the chapter *2.6.4.* Endogenous retrovirus genes and exogenous virus infection.

2.4.2.3 Python morulus endogenous retrovirus (PyERV)

PyERV probably causes a fatal disease - boid inclusion body disease - in boid snakes. Despite the clame that this retrovirus is unclassifiable, some place the virus in the genus Betaretrovirus. An endogenous copy of PyERV was named PyT2RV (Huder et al., 2002).

2.4.2.4 Pan troglotydes endogenous retroviruses (PtERVs)

PtERVs are a LTR retrotransposons present in three classes: CERV (Chimpanzee endogenous retrovirus) – gammaretroviruses, CERV II - betaretroviruses, CERV III -

spumaviruses. Majority of the chimpanzee-specific insertions belong to classes I and II. Frameshifts and substitutions damaged the ORFs of the proviruses so no functional gene could be expressed but eight copies could be retrotransposition-competent. Overall, PtERVs could drive genomic changes after the divergence of chimpanzees and humans (Mun, Lee, Kim, Kim, & Han, 2014).

Using viral constructs with MLV cores and reconstructed Envs of extinct CERV, the receptor of the virus was detected as the copper transporter (CTR1). The presence of the receptor on human germline cells does not support the fact that the presence enables endogenization, for no endogenous PtERVs or related sequences were detected in human genomes (Soll, Neil, & Bieniasz, 2010).

2.4.3 Gammaretroviruses

The presence of a Gammaretrovirus in any genome mostly does not occur to one as a surprise. The proportion of endogenous gammaretroviruses in host genomes is illustrated in the Figure 5. Since endogenous gammaretroviruses are abundant, one could expect to be more likely to come across a Gammaretrovirus that might be currently endogenizing. Despite the expectations resulting in presence of polymorphic integration in some genomes, the opposite is true. To thoroughly examine the process of endogenization, one needs a population with the similar integration pattern within it as well as the pattern being different to a one described elsewhere. This situation was so far described in the population of Australian koalas and mule deer inhabiting Northern part of America (discussed in previous chapters). The species in the genus are: HERVs - Human endogenous retroviruses, PERV - Porcine endogenous retrovirus, GALV - Gibbon ape leukemia virus, FeLV - Feline leukemia virus, MLV - Murine leukemia virus, KoRV - Koala retrovirus, CrERV - Cervid endogenous retrovirus.

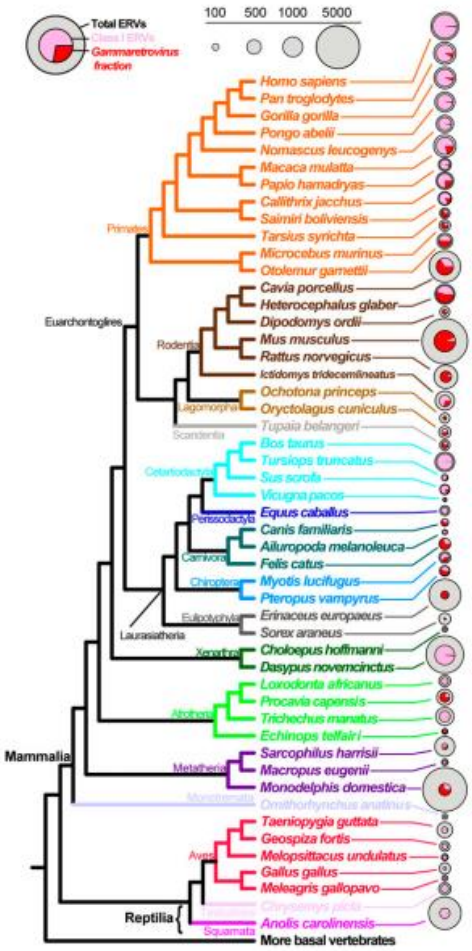

Figure 5: Endogenous gammaretroviruses among vertebrate genomes. The number of ERVs is depicted by the size of the outter gray circle. The proportion Class I ERV (Gamma- and Epsilonretrovirus) is depicted by inner circle. The proportion of gammaretroviruses is depicted by red in the inner circle (pie chart). The figure is adapted from (Hayward, Grabherr, & Jern, 2013).

2.4.3.1 Cervid endogenous retrovirus (CrERV)

The studies of CrERV dates back to the discovery of the first type C Gammaretrovirus discovered in the mammalian species of the New World origin (Aaronson et al., 1976). The discovered virus was denoted Deer kidney virus (DKV). The DKV virus was obtained by cocultivation with susceptible cells.

However, the virus became to be of higher interest when it was proved to be highly polymorphic and subsequently integration sites specific for a host population were described (Elleder et al., 2012; Kamath et al., 2014). Currently CrERV is a virus with one of the most polymorphic integration sites, whereas some patterns of integrations cluster geographically. This indicates that the endogenous copies of the retrovirus are

evolutionarily young and therefore the virus is a suitable model for studying the processes accompanying the early steps of endogenization. Despite the fact that the virus is efficient in creating new endogenous copies, it is inefficient as an exogenous virus in the original host species (discussed in this thesis).

2.4.3.2 Gibbon ape leukemia virus (GALV)

GaLV together with WMV (Wooly monkey virus) were the first exogenous retroviruses associated with leukemia in primates (Kawakami, Kollias, & Holmberg, 1980). Both of the viruses share a high sequence similarity together with KoRV (koala retrovirus). An endogenous copy of GaLV was detected in a rodent *Melomys burtoni* (Alfano et al., 2016). This finding might help explain the close relatedness of GaLV and KoRV for gibbons and koalas are not even geographically overlapping and the transfer of the retrovirus used to be rather a mystery.

2.4.3.3 Feline endogenous leukemia virus (FELV)

FeLV is one of the most studied endogenous retroviruses due to being the main caus of leucosis occurring in domestic cats which is their most common form of malignancy (Priester & Mantel, 1971). However, FeLV is not a threat to several populations of domestic cats only, but it also occurred in captive Asian leopard (Rasheed & Gardner, 1981), wildcat (Felis silvestris) (Boid et al., 1991), captive bobcat (Felis rufus) (Sleeman, Keane, Johnson, Brown, & Woude, 2001), in captive cheetah (Acinonyx jubatus) (Marker, Munson, Basson, & Quackenbush, 2003), but not in free ranging (Munson et al., 2004), panthers (Nolen, 2004), Florida pumas (Puma concolor coryi)(Cunningham et al., 2008), Iberian lynx (Lynx pardinus)(Meli et al., 2009), Pallas' cats (Felis manul) (Naidenko, Pavlova, & Kirilyuk, 2014), and guignas (Leopardus guigna) (Mora, Napolitano, Ortega, Poulin, & Pizarro-Lucero, 2015).

Being a threat to domestic cats and endangered species, a broad research concerning this pathogen has been conducted. Four FeLV subgroups (FeLV-A, FeLV-B, FeLV-C, and FeLV-T) were described. Subgroup B (which presence is an indicator of a poorer prognosis of leukemia (Sheets, Pandey, Jen, & Roy-Burman, 1993)) and C (causing non-fatal anemia (Mackey, Jarrett, Jarrett, & Laird, 1975)) are generated from subgroup A. Subgroup T is associated with immunosuppressive disease (Donahue et al., 1991).

2.4.3.4 Koala retrovirus (KoRV)

Leukemic and lymphoid neoplasia was proven to be present in koalas a long time ago. Findings in the year 1988 of gammaretroviral particles in koalas indicated that the pathologies might have a retroviral etiology (Canfield et al., 1988). Later the virus sequence was discovered to be similar to the sequence of GALV. Despite the high sequence similarity, the geographical distribution of koalas and gibbons makes the direct transmission among the two species highly improbable. However, these two viruses might have a common ancestor (Melomys burtoni retrovirus- MbRV) (Simmons et al., 2014).

2.4.3.5 Melomys burtoni retrovirus (MbRV)

MbRV was discovered by screening 42 either native or introduced species to Australia for the presence of KoRV-like retrovirus. The viral genome was sequenced. The sequence of *pol* and *env* cluster with KoRV and GaLV in the phylogeny trees. The presence of the virus particles was proven by electron microscopy, but the virus probably does not cause any cytopathic effect (Simmons et al., 2014).

2.4.3.6 Murine leukemia virus (MLV)

The MLV genome was the starting material in vector constructions in gene therapy. It is used as a model system in studies analyzing retroviral integration preferences. It may be due to the fact that it has a simple, well described genome (reviewed in (Rein, 2011)) and its subtypes are able to show different tropism: amphotropic - infecting all species, xenotropic - infecting only species different to the original host, and ecotropic - being able to infect only the original host. MLV is present in mice in endogenous form in many copies and some polymorphic integrations were described, but not in sufficient numbers to be utilized as a model for studying processes accompanying endogenization.

2.4.3.7 Porcine endogenous retrovirus (PERV)

PERV is considered a threat when it comes to xenotransplantation. Xenotransplantation has been performed in the passing of the history of medicine, mainly due to the lack of knowledge about the immunological interspecies barrier. However, using organs of other species for transplantation is still a current topic and some success was reported transplanting pig liver at least for the prolonged waiting time

for the human donor organ (Makowka et al., 1994). Still, the presence of PERV in the pig genome and possible zoonotic infection is still a major obstacle to overcome when it comes to xenotransplantation (reviewed in (Mattiuzzo, Takeuchi, & Scobie, 2012)). One of the most promising approaches towards this problem is the use of CRISPR-Cas technology to eliminate the presence of PERV in the pig genome (Yang et al., 2015). Due to its clinical significance, PERV is one of the best described endogenous retroviruses.

2.4.4 Deltaretroviruses

Deltaretroviruses are possibly the most mysterious group of retroviruses. An endogenous copy of Deltaretrovirus was not found for a long time. The belief that deltaretroviruses are not capable of creating endogenous copies was supported by the general belief that complex retroviruses are not efficient in infecting germline cells. However, we were successful in identifying the first presence of an endogenous deltaretroviral sequence.

As mentioned before, deltaretroviruses possess complex genomes coding for *gag-pol-env* and the sequences of additional genes. The additional genes for deltaretroviruses are Tax, Rex, and HBZ. Tax is an activator of viral and cellular transcription; Rex binds and stabilizes viral RNA, and HBZ plays a role in leukemogenesis and has multiple other functions. HBZ has the ORF of the sequence in the opposite strand compared to the rest of the provirus sequence. The species in this genus are: HTLV-1,2,3,4 - Human T-lymphotropic virus 1-4; STLV-1,2,3,4 - Simian T-lymphotropic virus 1-4; and BLV- Bovine leukaemia virus.

2.4.5 Epsilonretroviruses

Epsilonretroviruses are an exception for despite the fact they have complex genomes; they are able to form endogenous copies quite efficiently (see Figure 5 for reference). Epsilonretroviruses is the newest genus of Orthoretrovirinae. Endogenous copies of these viruses were found in fish and amphibians.

Walleye dermal sarcoma virus (WDSV) causes dermal sarcomas in its piscine host - walleye (*Stizostedion vitreum*). The accessory genes regulate the host metabolism and induce cancerogenous changes. The most distinc feature of the cancerogenesity is its seasonal cycle (Bowser, Wolfe, Forney, & Wooster, 1988) and a complex life cycle with

varying gene expression pattern throughout the replication of the virus and stage of the disease (Quackenbush, Holzschu, Bowser, & Casey, 1997).

Walleye epidermal hyperplasia virus (WEHV) was found in two subtypes: WEHV1 and WEHV2. Both are probably causing hyperplasia in walleye. Similarly to WDSV, the occurrence of the pathology has seasonal cycles (LaPierre, Holzschu, Wooster, Bowser, & Casey, 1998).

Xenopus laevis endogenous retrovirus (Xen1) with length over 10 kb is probably of the largest endogenous retrovirus known. It has only four frameshift mutations and no obvious stop codon. The 99% similarity of the provirus LTRs suggests that the virus integrated into the genome recently (Kambol, Kabat, & Tristem, 2003).

2.4.6 Lentiviruses

Lentiviruses were thought to be a young genus of retroviruses for their endogenous copy was not detected for a long time. Due to the quite recent HIV outbreak, lentiviruses might be currently the most studied retroviruses. They cause severe pathologies. The members of this group with the important notion are HIV (human immunodeficiency virus) and other immunodeficiency viruses such as of simian species (SIV) and felids (FIV). Lentiviruses possess complex genomes. Besides *gag*, *pol*, and *env*, they also bear additional accessory genes.

Discoveries of their endogenous copies pushed the knowledge about their evolution several million years deeper. These discoveries also provide an interesting insight about host-interaction evolution via combining information about accessory genes and analysis of evolution of their counteracting restriction factors.

2.4.6.1 Lentiviral regulatory and accessory genes

Not all of the lentiviruses bear all of the regulatory and accessory genes. Based on the discoveries of their endogenous forms we are able to study the evolution of lentiviral accessory genes as well. The evolutionary tree indicating the evolutionary dynamics of presence/absence of some accessory genes is depicted in the Figure 7.

2.4.6.1.1 Regulatory genes

2.4.6.1.1.1 Tat

Tat acts as a trans-activator during transcription to enhance initiation and elongation. Tat binds nascent RNA with TAR (Tat-responsive element) in LTR. It works in cooperation with cellular protein pTEFb which binds Tat and TAR. pTEFb phosphorylates polymerase II and thus increases the processivity of the polymerase (reviewed in (Zou, Peng, Wang, & Zhou, 2016)). Tat is not present in feline lentiviruses, but many binding sites for enhancer proteins are present in LTRs of felid lentiviruses. However, their presence was not proven to be essential for replication of felid lentiviruses (Miyazawa, Tomonaga, Kawaguchi, & Mikami, 1994).

2.4.6.1.1.2 Rev

The abbreviation Rev stands for "regulation of expression of virion proteins". Rev contains and arginine-rich RNA binding domain which binds to RRE (Rev responsive element). The Rev responsive element acts post-transcriptionally, regulating mRNA splicing and transport to the cytoplasm. It works with cellular proteins binding Rev (e.g. importinβ). Rev contains NLS (nuclear localization signal) which aids its return to the nucleus (reviewed in (Grewe & Uberla, 2010). Presence of Rev was not detected in endogenous Lentivirus of Malayan colugo (Han & Worobey, 2015; Hron et al., 2014). As discussed previously, the accessory gene might have been present in the circulating retrovirus at the time of infection, but due to being heavily mutated, we are not able to detect it, so our presumption of its absence might be biased. However, if Rev was lacking in the genome of the virus which integrated into the colugo genome, the virus might have been using different replication strategy than current lentiviruses. It was proven that viruses with impaired Rev exhibit faulty replication in various steps (Blissenbach, Grewe, Hoffmann, Brandt, & Uberla, 2010).

2.4.6.1.1.3 P6

P6 is technically not a regulatory gene, but is included here for if it is disrupted, the lentiviruses coding for P6 in its genome are budding only with difficulties and therefore its evolutionary study might give an additional insight on host-Lentivirus interactions. P6 is a proline-rich protein interacting with endosomal vesicles. However, besides studying it in HIV and primate lentiviruses (Bibollet-Ruche et al., 2004), not

much attention is paid to it. It was not described in any of the discovered endogenous lentiviruses.

2.4.6.1.2 Other accessory genes

When it comes to identification of accessory genes in endogenous retroviruses, two possible scenarios if the accessory gene is stated to be lacking might occur. Either the area bearing was already heavily mutated and hence the presence of an accessory gene cannot be determined properly or the accessory gene was lacking in the virus that endogenized. This leads to contradictions presented in the works studying lentiviral accessory genes.

2.4.6.1.2.1 Vif

The abbreviation Vif stands for virus infectivity factor. Vif aids virion maturation and infectivity. Vif is not present in equine lentiviruses and was not detected in the endogenous Lentivirus of colugo (Han & Worobey, 2015). Vif is stated to be present in RELIK by some (Han & Worobey, 2015) and stated to be lacking by others (Katzourakis et al., 2007). However Vif is detected in the genome of ELVmpf (endogenous Lentivirus of Mustela putorius furo) which happens to be dated as an older lentiviral lineage. Therefore the presence of Vif in the retrovirus endogenizing as RELIK occurs as more probable.

2.4.6.1.2.2 Nef

Nef is also known as a negative factor for originally it was described as a redundant accessory gene. However, it was proven to aid pathogenicity of the virus and if deleted from the SIV genome, it reduces pathogenicity in macaques (reviewed in (Laguette, Bregnard, Benichou, & Basmaciogullari, 2010). It was proposed that infection by Nef-lackig SIV might protect macaques from infection by more aggressive SIV forms. Nef was probably acquired only by simian and human immunodeficiency viruses.

2.4.6.1.2.3 Vpu

Vpu stands for “Viral protein-unknown”. Vpu downregulates CD4 in cells and increases virion release from the cells. The gene coding for Vpu was detected only in HIV-1 and HIV-1 related SIV isolates. HIV-2 and majority of the SIV isolates lack this accessory gene (Hussain, Wesley, Khalid, Chaudhry, & Jameel, 2008). It was proven to be

capable of counteracting the cellular restriction factor Tetherin. However, the HIV-1 group O uses Nef to counteract the action of Tetherin (Bush & Tebit, 2015). This general absence of Vpu in the lentiviral group might indicate that this is rather a young accessory gene.

2.4.6.1.2.4 Vpr (Vpx)

Vpr stands for "Viral protein-regulatory". Vpr induces G2 arrest of the cell cycle and induces apoptosis. Vpr is encoded by HIV-1. HIV-2, SIVsm, and SIVmac encode two proteins that are homologous to HIV-1 Vpr, namely Vpr and Vpx (reviewed in (Planelles & Barker, 2010). Vpx is encoded by HIV-2 and simian immunodeficiency viruses. Vpx counteracts the actions of SAMHD1 (Herrmann, Happel, & Gramberg, 2016). As with Vpu, the presence of this accessory gene only in human and simian immunodeficiency viruses indicates its rather recent acquisition by lentiviral genomes.

2.4.6.1.2.5 dUTPase

dUTPase is technically not an accessory gene. It is not a solitary gene, but is encoded by the *pro* gene or part of *pol* gene in some viruses. However, I discuss it here for its interesting evolutionary dynamics among lentiviruses. What comes as a surprise is the fact that dUTPase is encoded in various places in viral genomes (see Figure 6). This suggests that viruses might have acquired dUTPase during various evolutionary events (Hizi & Herzig, 2015).

dUTPase is present in the genomes of EIAV (Equine infectious anemia virus), FIV (Feline immunodeficiency virus), and CAEV (Caprine arthritis encephalitis virus), but not in primate lentiviruses. The function of dUTPase in lentiviral genomes is still not clearly elucidated. EIAV and FIV with removed dUTPase replicate poorly in the cultured macrophages and viral load or severity of symptoms might be decreased in *in vivo* infections (Lerner et al., 1995; Lichtenstein et al., 1995).

However, when it was removed from the CAEV, the virus was shown to replicate slowly in macrophages, but after some time it replicated at the same rate as wild-type CAEV. CAEV without dUTPase differs in pathogenesis compared to virus with intact dUTPase. The presence of dUTPase was also shown to increase genomic stability of CAEV by preventing G-to-A mutations (Turelli, Guiguen, Mornex, Vigne, & Querat, 1997).

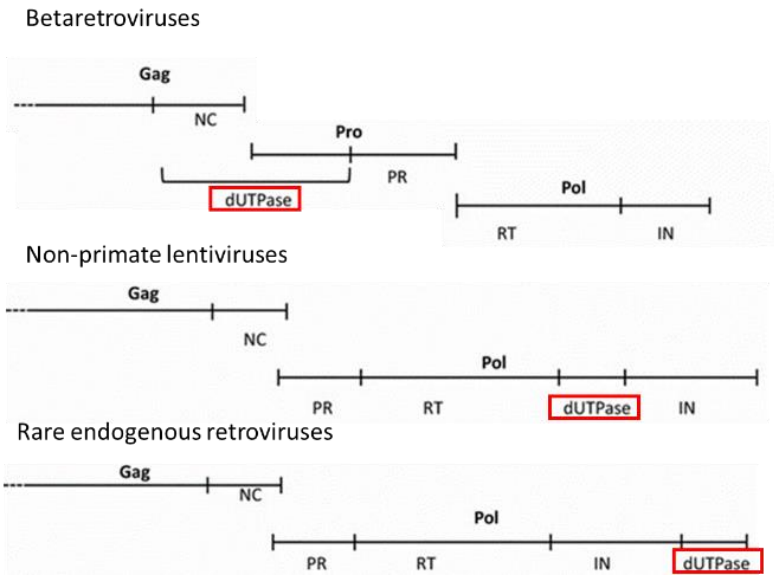

Figure 6: Position of dUTPase in retroviral genomes. The schemes are not drawn to scale. The figure is adapted from (Hizi & Herzig, 2015).

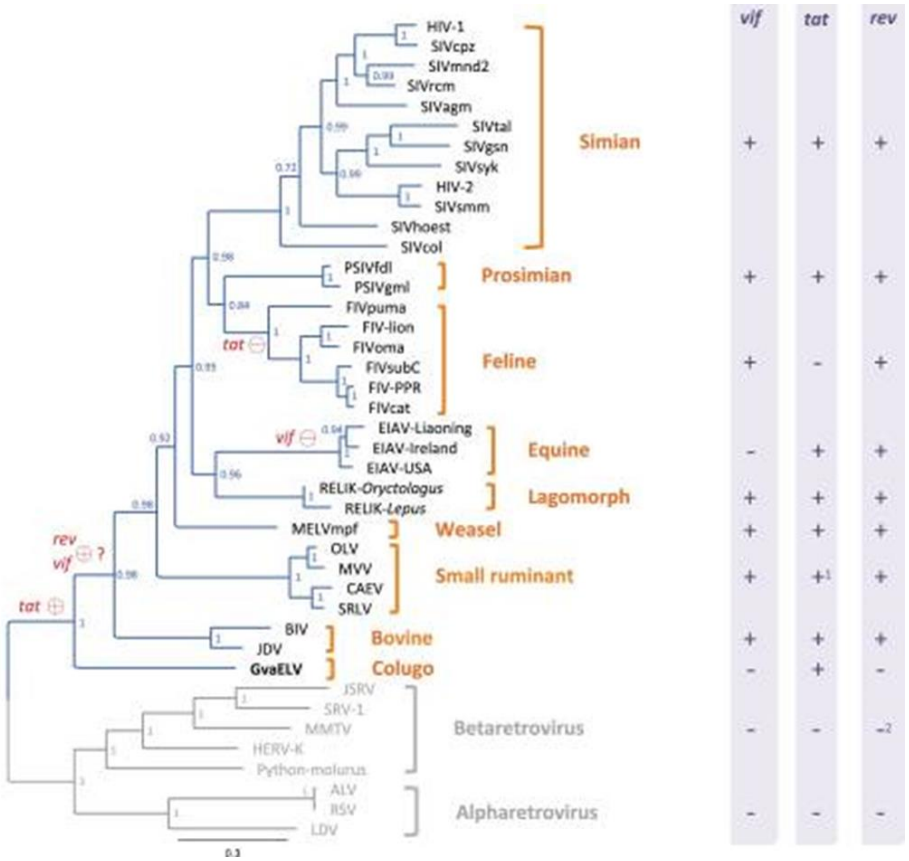

Figure 7: Presence/absence of the accessory genes. Adapted from (Han & Worobey, 2015).

2.4.6.2 Endogenous lentiviruses

Lentiviruses were believed not to form endogenous counterparts for quite a long time. Firstly, they were believed to be evolutionary young and secondly, they are complex retroviruses which were thought not to be able to form an endogenous copy at all. This was hypothesized to be either due to lack of receptor on germline cells or due to virus cytopathic effects. This belief was proven to be wrong. Endogenous Lentivirus was first discovered in the rabbit genome (Katzourakis et al., 2007). Subsequently, endogenous lentiviruses were discovered in lemurs (Keckesova, Ylinen, Towers, Gifford, & Katzourakis, 2009) and ferrets (Cui & Holmes, 2012a).

2.4.6.2.1 Endogenous Lentivirus of Mustela putorius furo (ELVmpf)

ELVmpf was the third endogenous Lentivirus discovered. Besides lentiviral *gag-pol-env* genes, Vif-like element was detected. The insertion into the ferret genome was estimated to occur 12 MYA (Cui & Holmes, 2012a). The further study of ELVmpf confirmed its presence in the species of *Lutrinae* and *Mustelinae* subfamilies but not the *Martinae* subfamily. This confirmed the estimated age of the provirus. An additional accessory gene was identified in endogenous lentiviruses- *vif* (Han & Worobey, 2012). It was proposed that ferrets could be used as a model to study lentiviral-host interactions for their cells can be productively infected by HIV-1 (Fadel et al., 2012)

2.4.6.2.2 Endogenous simian immunodeficiency virus (pSIV)

An endogenous form of simian immunodeficiency virus was detected in the WGS (whole genome shotgun) data of Grey mouse lemur (*Microcebus murinus*) and denoted pSIVgml. The presence of pSIVgml was subsequently detected in samples from six additional species of lemur. The invasion of the genome was estimated to have occurred 4 MYA. This estimation is still an object to be questioned, for pSIV is present both in lemurs inhabiting Madagascar and African landmass. These two areas are divided by a 400 km wide and deep ocean for the last 130 million years, creating a geographical barrier hard to be crossed by lemurs. pSIV is the first and so far only evidence of an endogenous Lentivirus presence in a primate genome. *Vpr* and possibly *nef* sequences are present in the proviral sequence, but dUTPase seems to be lacking (Gifford et al., 2008).

2.4.6.2.3 Rabbit endogenous Lentivirus type K (RELIK)

The first endogenous Lentivirus was discovered in the genome of European rabbit (*Oryctolagus cuniculus*). The provirus bears full-length *gag, pol*, and *env* genes, but highly mutated (frameshifts and stop codons included). The endogenous copy posseses an element resembling genes for *tat* and *rev*. The presence of relatively intact endogenous copies and several solo LTRs point out the established germline infection occuring approximately 5 MYA (Katzourakis et al., 2007).

The CA of RELIK was proven to interact with Trim5α. Trim5α was also proven to be under positive selection in Leporidae (Yap & Stoye, 2013). These facts might indicate the relationship between Trim5α and RELIK, but it is very hard to assign the presence of a positive selection towards a restriction factor to a specific virus. The CA of RELIK was also proven to interact with cyclophilinA (CypA). CypA was packaged into virions containing ancient lentiviral CA. The complex CA-CypA of ancient proviruses probably enabled the viruses to infect nondiving cells, protected them from restriction factors or play a role in the virion assembly or CA uncoating (Goldstone et al., 2010).

2.4.7 Spumaviruses

Placed outside of the Orthoretroviral clade, spumaviruses make up a special retroviral group. Due to the fact they they do not cause severe pathologies, they were not of a deeper interest for a long time. This feature became lately an advantage for it makes them a great candidate to be used in gene therapy. As exogenous, they infect broad range of animals. Some of the species in the genus are: African green monkey simian foamy virus, Macaque simian foamy virus, Bovine foamy virus, Equine foamy virus, Feline foamy virus.

2.4.7.1 Sloth endogenous foamy virus (SloEFV)

The presence of approximately 11,5 kb long SloEFV is the evidence that foamy viruses could be infecting ancestral mammals more than 100 MYA (Katzourakis, Gifford, Tristem, Gilbert, & Pybus, 2009). Despite the fact that SloEFV was described for some years, not much attention was paid to it nor was its more recent evolution studied. This might be partly owed to the fact that not many extant relative host species exist. The most recent work analyzes the ancient sequence found in modern sloths and 13,000 and

20,000 years old samples. Lineage-specific SloEFV copies were detected and the age of the repeated recent endogenization (10-20 MYA) was proposed (Slater et al., 2016).

2.5 Retroviral life cycle and the host

When it comes to discussing any pathogen and the host relationship, the Red Queen Hypothesis proposed by Van Valen always comes to notice. Retroviruses are no exception. Since the hosts and retroviruses co-existed for millions of years, their relationship naturally evolved as well.

2.5.1 Receptor block

The receptor is the virus's gateway to the cell, making it the first line of defense against the virus for the cell. As mentioned before, retroviruses usually exploit cellular receptors used for physiologically occurring transport of various substances to the cell. However, the presence or absence of a particular receptor does not rule out nor indicate the potential of a retrovirus to endogenize. Some viruses utilize the same receptor, whereas some subtypes of a retrovirus utilize different receptors (see Table 2 for review). The occurrence of the replication block at a receptor lever might be determined by changed conformation of receptors or co-receptors.

The other possible scenarios for the occurrence of receptor block are the changes in viral envelope. Of all the genes of retroviruses, *env* probably displays the most variable functionality and possess the most rapidly evolving sequence. The major function is the binding to the receptor. Some additional activities were detected in various retroviral Envs. For example, in MuLV, Env was proven to affect membrane fusion. JSRV Env transforms cells *in vitro* and work as an oncogene *in vivo*. The C-terminal tail of Env in lentiviruses is uncommonly long and besides determining Env structutal and functional features, plays a role in the maturation of virions and formation of late endosomes (reviewed in (Steckbeck, Kuhlmann, & Montelaro, 2014)).

Table 2: Retrovirus receptors

Retrovirus	Receptor	Reference
HIV, SIV	CD4 helper T cell receptor, CXCR4, CCR5, and others	(Dalgleish et al., 1984) (Klatzmann et al., 1984) (Maddon et al., 1986)
MLV-E	CAT-1 (SLC7A1) (Specific membrane receptor expressed on mouse cells)	(Albritton, Tseng, Scadden, & Cunningham, 1989)
MLV-A	Ram-1 and GLVR-2 (Cellular receptor for amphotropic murine retroviruses)	(Miller, Edwards, & Miller, 1994) (van Zeijl et al., 1994)
GALV	GLVR-1	(O'Hara et al., 1990)
GALV, 10A1 MLV, FeLV-B, woolly monkey virus	Pit1 (SLC20A1) (phosphate transport)	Reviewed in (Overbaugh, Miller, & Eiden, 2001)
A-MLV, 10A1 MLV, FeLV-B, BLV	Pit2 (SLC20A2) (phosphate transport)	Reviewed in (Overbaugh et al., 2001)
Xenotropic and Polytropic MLVs	XPR1 (G-protein coupled signaling)	Reviewed in (Overbaugh et al., 2001)
MMTV	Mtvr, Protein of unknown function	(Stewart, 2002)
FeLV-B	The same receptor as GALV	(Takeuchi et al., 1992)
FeLV-C	Flvcr (anion transporter)	(Quigley et al., 2000)
FeLV-T	FeLIX and Pit1 (SLC20A1); Env like protein	(Anderson, Lauring, Burns, & Overbaugh, 2000)
ASLV-A	*tva*-member of LDL receptor family	(Gilbert, Bates, Varmus, & White, 1994)
ASLV-C	*tvc* (member of immunoglobulin superfamily)	(Elleder, Plachy, Hejnar, Geryk, & Svoboda, 2004)
ASLV-B	Various alleles of *tvb* (member of TNF receptor superfamily)	(Smith, Brojatsch, Naughton, & Young, 1998)
ASLV-D		(Smith et al., 1998)
ASLV-E		(Klucking, Adkins, & Young, 2002)
RD-114, type D SRV, BaEV, HERV-W	*RDR(SLC1A5) or RDR2(SLC1A4) (neutral amino acid transport)*	Reviewed in (Overbaugh et al., 2001)
BLV	Blvr reported by (Ban et al., 1993), was reported to be wrongly identified; the receptor is cationic amino acid transporter CAT1 (SLC7A1)	Reported at Cold Spring Harbor Retrovirusese meeting by Jean-Luc Battini
JSRV	HYAL2 (hyaluronidase)	(Miller, 2008)

2.5.2 Later replication blocks

The block of the retrovirus replication might occur later in the retrovirus life cycle. After the entrance into the cell, the efficient disassembly of the viral particle, formation of pre-integration complex, reverse transcription or integration might be restricted (early blocks). The transcription of the integrated provirus as well as translation, efficient assembly, release or maturation of the viral particle might be restricted as well (late blocks). Cells usually possess mechanisms to block various step of the retroviral life cycle. Some of them might be mediated by intracellular host restriction factors.

2.5.3 Intracellular restriction factors

Restriction factors are a part of the innate immune system of the host cell. These intrinsic proteinaceous antiviral immune effectors are often induced by IFNα or antiviral activity. In complex viruses, viral antagonists for specific restriction factors are found. They do not share a specific trait, such as a sequence motif or specific structure. The list of restriction factors described deals with the most studied ones in relationship with retroviruses.

2.5.3.1 APOBECs and AID family

The abbreviation APOBEC stands for apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like. AID encoded by *aicda* gene stands for Activation-induced cytidine deaminase. APOBECs are a family of evolutionary young proteins catalyzing cytidine deamination (C to U editing reaction). APOBECs are present in the vertebrate genomes in many subtypes (up to 11), whereas present in only one in sub-vertebrates. APOBECs 3 are supposed to be under positive selection (based on dN/dS ratio) and probably occurred as an APOBEC1/AID duplication (reviewed in (Harris & Dudley, 2015)).

APOBECs are widely studied in relationship with HIV. The sequences are assembled only in some ungulate species and well annotated only in *Bos taurus* and *Ovis aries*. The APOBECs functions described in human are listed in Table 3.

Table 3: The individual APOBEC and AID functions.

	Function	Reference
APOBEC1	Cholesterol metabolism, oncogene, viral restriction	(Rosenberg, Hamilton, Mwangi, Dewell, & Papavasiliou, 2011)
APOBEC2	Muscle and heart specific APOBEC	(Liao et al., 1999)
APOBEC3A	Cytosine deamination of foreign DNA	(Stenglein, Burns, Li, Lengyel, & Harris, 2010)
APOBEC3B **APOBEC3C** **APOBEC3E**	RNA editing, cell cycle control	(Jarmuz et al., 2002)
APOBEC3F	RNA editing, cell cycle control Could inhibit accumulation of HIV-1 RT products	(Jarmuz et al., 2002) (Holmes, Koning, Bishop, & Malim, 2007)
APOBEC3G	Affects HIV replication in various steps: RT- Inhibits priming of tRNA and thus production of viral ssDNA Integration- APOBEC3 can cause aberrant LTRs formation, incapable of integration into the host genome Have been also shown to inhibit: gammaretroviruses, deltaretroviruses, lentiviruses spumaviruses, and retrotransposition of endogenous murine gammaretroviruses, mainly by cytosine deamination of foreign DNA	(Jarmuz et al., 2002) (Guo, Cen, Niu, Saadatmand, & Kleiman, 2006) (Mbisa et al., 2007) (Doehle, Schafer, Wiegand, Bogerd, & Cullen, 2005) (Sasada et al., 2005) (Lochelt et al., 2005) (Esnault et al., 2005)
APOBEC3H	Antiviral function in old world monkeys, but suppressed in humans	(OhAinle, Kerns, Malik, & Emerman, 2006)
APOBEC4	Lacks cytidine deaminase activity; boosts promoter activity and HIV replication.	(Marino et al., 2016)
AID	Expressed in B-cells. Immunoglobulin gene class switches DNA recombination and somatic hypermutation.	(Muramatsu et al., 2000)

2.5.3.2 Mx

Mx (Myxovirus resistance) is one of the interferon-inducible restriction factors. *Mx1* codes Mx1 in mice and MxA in humans. Myxovirus A resistance protein (MxA) is elevated in the blood after the viral infection independent of the virus species, but is mostly studied in relationship with Orthomyxoviruses. Mx2 (also known as MxB) was

shown to inhibit HIV-1 and similar Lentivirus infections (Goujon et al., 2013; Kane et al., 2013; Liu et al., 2013).

Mx2 was shown to inhibit the HIV-1 nuclear entry (Goujon et al., 2013). Mx2 was also shown to interact with HIV-1 capsid and mediate the block of retroviral replication early after virus enters the cell. Cyclophilin might play a role in mediating the restriction in other HIV-1 life cycle stages as well (Bulli et al., 2016)

2.5.3.3 SAMHD1

SAMHD1 (Sterile alpha motif and histidine-aspartate domain containing protein 1) is a restriction factor that depletes the dNTP pool in the cell to block retroviral replication by restriction of reverse transcription. It is counteracted by the accessory HIV gene product Vpx. The mutations in the SAMHD1 gene cause Aicardi-Goutières syndrome which is a genetic encephalopathy mimicking congenital viral infection (Powell, Holland, Hollis, & Perrino, 2011).

This indicates that SAMHD1 probably plays a role in regulating innate immune response. The counteracting viral gene product Vpx is present only in HIV-2 and some SIV strains making these viruses more resistant to SAMHD1 restriction via unique clade specific SAMHD1-Vpx interactions (Wu et al., 2015).

SAMHD1 has enzymatic activities such as acting as a dNTPase and nuclease as well as being able to bind single-stranded DNA/RNA. The precise molecular action of SAMHD1 is still not completely elucidated despite thorough studies of the protein, including the structural ones (reviewed in (Ahn, 2016)).

2.5.3.4 TREX1

Formerly known as DNaseIII, TREX1 (three prime repair exonuclease 1) is the major 3’ to 5’ DNA-specific exonuclease in mammalian cells (Hoss et al., 1999). The mutation in the TREX1 gene are associated with autoimmune diseases (reviewed in (Rice, Rodero, & Crow, 2015)). TREX1 plays a role in HIV infection via inhibiting cytosolic DNA sensing pathway and thus modifying the interferon response to infection as well (reviewed in(Hasan & Yan, 2014)).

2.5.3.5 TRIM5α

TRIM5α (Tripartite motif alpha) is a member of one of the eleven subgroups of TRIM proteins (see Figure 8 for reference). TRIMs are intrinsic immunity factors and have direct antiviral activity. Various TRIMs have been proven to act against different viruses (reviewed in (Nisole, Stoye, & Saib, 2005)). TRIM5 induces NFκB activation. TRIM5 recognizes the retroviral capsid by its PRY/SPRY domain. These facts indicate that TRIM5 as well as TRIM21 might be directly acting as pattern recognition receptors (Keeble, Khan, Forster, & James, 2008; Pertel et al., 2011). TRIM5α in macaques was proven to inhibit reverse transcription activity of HIV-1 as well as abort efficient disassembly of the viral particles (Campbell et al., 2015).

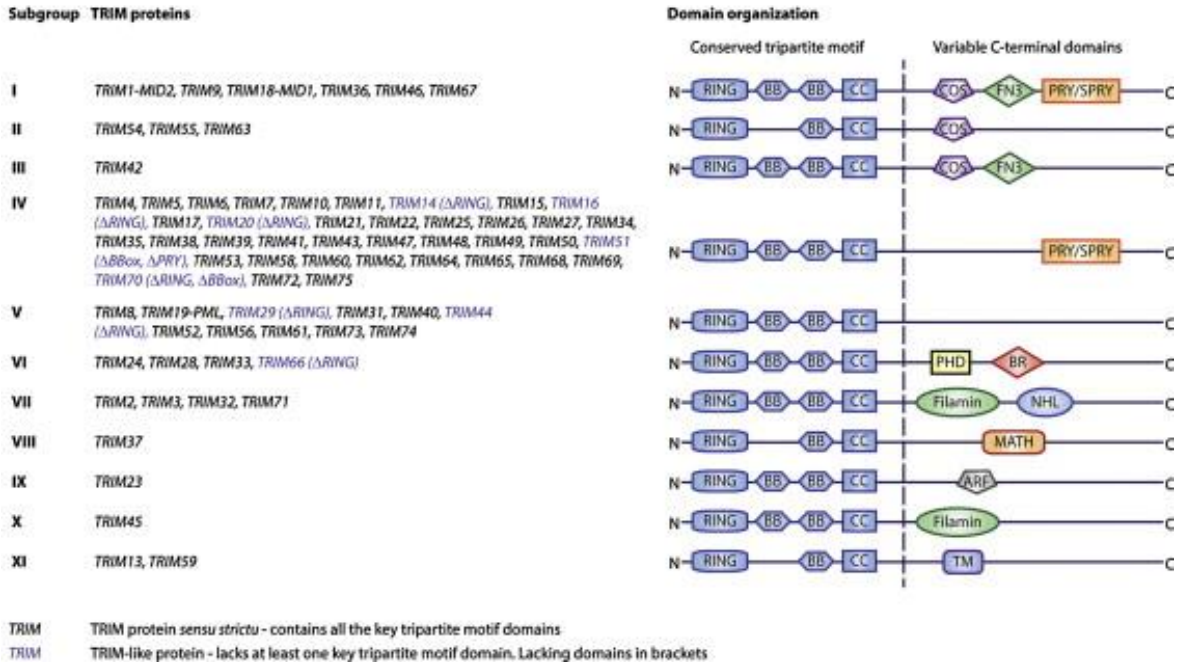

Figure 8: Domain organization of TRIMs (from (Versteeg, Benke, Garcia-Sastre, & Rajsbaum, 2014)).

2.5.3.6 Viperin

Viperin was not shown to play a role in fighting retroviral infections.

2.5.3.7 Tetherin

Tetherin was discovered as a protein expressed on the surface of human plasmatic cells in cell lines (Goto et al., 1994). The function of Tetherin is to tether the viral particles inside the host cells and therefore preventing the further spreading of

infection. The HIV-1 codes an accessory gene product Vpu which counteracts this restriction. However, when the Vpu is lacking, virions are retained at the cell surface and consequently endocytosed (Perez-Caballero et al., 2009). Despite the fact that Tetherin is known to be one corresponding gene present in one isoform in human, it is present in three isoforms in sheep (Arnaud et al., 2010) and in three isoforms in cattle (Takeda et al., 2012). Tetherins of various species were proven to inhibit PERV replication (Abe, Fukuma, Yoshikawa, Miyazawa, & Yasuda, 2014; Bae & Jung, 2014a, b; Mattiuzzo, Ivol, & Takeuchi, 2010; Mattiuzzo & Takeuchi, 2010).

2.6 Endogenous retrovirus presence in the host genome

The presence of an endogenous retrovirus in the host genomes might have various impacts. Besides the fact, that the presence of a retrovirus might be a cause of pathology, it might be an asset for the host. The endogenous retrovirus might either work as a restriction factor preventing other infections or even play a part in the normal physiology or cellular functions. This phenomenon is reviewed in this chapter. However, there are other mechanisms how the retroviral integration at a specific integration site might affect the host.

The insertion of the retrovirus into a gene may cause the gene disruption or gene up-regulation. Based on the site of integration, the virus can bring a promoter indefinitely inducing a transcription of a particular gene, leading to undesired pathology. The studies dealing with the integration site of a retrovirus (or a model construct representing one) often analyze not only the sequence of the integration site, but also the epigenetics of the chromosome area (e.g. methylation) of the site in order to estimate the probability of the retrovirus being transcribed.

There is another mechanism by which a virus can cause deregulation in the signaling pathway. The virus might code a gene from the pathway leading to overexpression of the gene while the virus genes are expressed. Some genes coded by various retroviruses that might deregulate the signaling pathways. The first described proto oncogene coded in a retrovirus genome was *src* in the genome of RSV (reviewed in (Martin, 2004)).

2.6.1 Endogenous retrovirus and the immunity of the host

Besides triggering the action of restriction factors, a new concept of role of endogenous retroviruses in innate immunity response was recently proposed in T-cell independent B-cell response. Multivalent molecules with repetitive structures trigger antibody response through B cell receptor crosslinking with T-cell help absent. This causes upregulation of endogenous retrovirus RNAs in antigen-specific mouse B cells which may be detected by MAVS (mitochondrial antiviral signaling protein) triggering the MAVS-dependent RNA sensing pathway. The RNAs might be transcribed in the presence of reverse transcriptase and trigger the cGAS-cGAMP-STING [cGAS (Cyclic GMP-AMP synthase) –cGAMP (Cyclic guanosine monophosphate–adenosine monophosphate) –STING (Stimulator of interferon genes)] pathway promoting the immunoglobulin M production. Deficiency of MAVS, cGAS or reverse transcriptase in the cell dramatically decreases the immunological response suggesting that endogenous retroviruses might play a role in the B-cell response (Zeng et al., 2014) .

2.6.2 Endogenous retrovirus genes and exogenous virus infection

The presence of retrovirus might have an impact on the host not only by affecting the genes surrounding its integration site or by delivering oncogenes to the host. It might be also utilized as a source of gene products included in the antiviral immunity. This concept was described as “Fighting fire with fire” by (Malfavon-Borja & Feschotte, 2015). On the contrary, products of endogenous retroviruses might be required for successful infection by an exogenous retrovirus. These cases are reviewed in the chapter 2.6.1.1.2 “Adding fuel to the fire”.

2.6.2.1 “Fighting fire with fire”

A longstanding concept that aberrant expression of retroviral genes might serve as a protection against further retroviral infection has been studied for several decades (Robinson, Astrin, Senior, & Salazar, 1981). Especially the proteins coded by the *env* (Malfavon-Borja & Feschotte, 2015) and *gag* (Mura et al., 2004) genes were shown to act as restriction factors in chicken, mice, sheep, and cats.

2.6.2.1.1 Fv1

Fv1 was first described as a gene determining the susceptibility of mice to various strains of MLV. The precise mechanism of Fv1 action is not known, but it was proved that it restricts the infection after reverse transcription and prior to integration (Jolicoeur & Baltimore, 1976). Fv1 shares the general structural features with TRIM5α. Both Fv1 and TRIM5α bind the retroviral capsid and restrict the virus in the similar manner. (Sanz-Ramos & Stoye, 2013). *Fv1* gene is present in multiple alleles. The levels of expression from individual alleles vary. Different alleles show restriction specificity, which is not generally affected by their expression level (Li, Yap, Voss, & Stoye, 2016).

Fv1 is sequentially homologous to *gag* gene of an endogenous retrovirus. Fv1 was found to be under positive selection in all analyzed *Mus* subgenera, consistent with its role in antiviral defense (Yan, Buckler-White, Wollenberg, & Kozak, 2009). The evolutionary analysis revealed that mice have probably acquired endogenous retrovirus giving rise to Fv1 approximately 5 MYA (Yap, Colbeck, Ellis, & Stoye, 2014).

2.6.2.1.2 Fv4/Akvr1

Fv4 and Akvr1 are alternative names for the same gene, which is an expressed *env* gene of an endogenous ecotropic MLV (Kozak, 2014). Fv4 restricts the infection by binding to the receptor for exogenous ecotropic MLV, however the restriction is not absolute (Takeda & Matano, 2007). A mild restriction effect on amphotropic MLV was proven as well.

2.6.2.1.3 Rmcf

Resistance to mink cell focus-forming virus (Rmcf) is a product of an endogenous polytropic MLV and protects against polytropic MLV infection. Rmcf structure resembles the one of Fv4. *Mus castaneus* was proven to lack endogenous retrovirus producing Rmcf. The interference mechanism was characterized in receptor restriction mediated by Rmcf (Jung, Lyu, Buckler-White, & Kozak, 2002).

2.6.2.1.4 Rmcf2

Rmcf2 is a protein similar to Fv4, becuase it is an expressed *env* of a xenotropic provirus present in *Mus castaneus* genome, unable to produce infectious particles. It protects *Mus castaneus* against polytropic MLVs through interference. *Rmcf2* is the third

described mouse gene of retroviral origin delivering protection against MLV infection (Wu, Yan, & Kozak, 2005).

2.6.2.1.5 enJSRV gag and JSRV

The Gag of enJSRV was the first endogenous retroviral Gag described as one with interfering properties to exogenous retrovirus infection by mediating late-stage replication block (Mura et al., 2004). This process is called JLR (JSRV late restriction), it became a model for studying late replication blocks in the retroviral life cycle. Because of the fact that JSRV is not able to grow in cell tissue culture (due to the lack of permissive cell line), the experiments performed after retroviral transfection studied the formation of the Gag aggregates and their disruptions (Arnaud, Murcia, & Palmarini, 2007b).

Still, there exists one exception to this restriction. An evolutionary young copy of JSRV (estimated as approximately 200 years old), denoted enJSRV26, is able to overcome the restriction mediated by endogenous *gag*. The counteraction of endogenous restriction is mediated by the *env* of enJSRV26. Amplification of the copies of enJSRV26 within the genome indicates that the sheep genome is still probably invaded by JSRV both endogenously and exogenously (Armezzani et al., 2011).

2.6.2.1.6 Refrex-1 and FeLV

Endogenous copies of FeLV were detected in the cat genomes. Exogenous feline retroviruses and feline ERVs have complicated genetic interactions: recombination between FeLV and ERV-DC (feline ERV) generated FeLV-D. Lately, a subgroup of FeLV-D was experimentally constructed combining an *env* of endogenous FeLV of domestic cats (ERV-DC) and exogenous FeLV. Close examination of the construct led to the discovery of the endogenous restriction factor Refrex-1 originating from *env* of ERV-DC (Ito et al., 2013). By reverse mutation, functional *env* was reconstructed from ERV-DC. This “reverse evolution experiment” indicates, that *env* of FeLV was repeatedly inactivated (Ito, Baba, Kawasaki, & Nishigaki, 2015).

2.6.2.1.7 CHOK1 cells and amphotropic retroviruses

Chinese hamster ovary cells (CHOK1) harbor a variety of endogenous retroviruses. Chemical activation of CHOK1 cells with 5-Bromodexouridine induced

production of atypical retroviral C-type particles (Manly, Givens, Taber, & Zeigel, 1978). CHOK1 cells are not susceptible to infection by exogenous retroviruses or retroviral vectors with an amphotropic MLV envelope. CHOK1 cells are widely used for biotechnology applications and amphotropic viruses are commonly used as vectors. Introduction of human amphotropic or mouse ecotropic retrovirus receptor or tunicamycin treatment of the cells (deglycosylation of the receptor) was shown to rescue the aforementioned susceptibility (Miller & Miller, 1992). However, no detailed characterization and explanation of the resistance mechanism was available.

2.6.2.2 "Adding fuel to the fire"

As reviewed in the previous chapter, products of the endogenous retroviruses can serve as endogenous restriction factors, preventing infection by exogenous viruses. On the other hand, some remnants of endogenous retroviruses might act as enhancers of retroviral infection or can even be required for the infection to proceed. These products of endogenous retroviruses are reviewed in this chapter.

2.6.2.2.1 FeLIX and FeLV

Besides multiple membrane-spanning receptor molecules (e.g. Pit1), the subgroup T of FeLV (FeLV-T) requires a cellular cofactor FeLIX for productive infection. FeLIX is endogenously expressed and is sequentially similar to the *env* of FeLV (Anderson et al., 2000). Several sequences of endogenous FeLVs were detected in cats, but none of them was genetically fixed. However, FeLIX activity was detected in sufficient amounts in all of the tested cats from various domestic regions. This finding indicates that the FeLV capable of producing FeLIX entered the cat genome prior to the FeLVs recently isolated (Sakaguchi, Shojima, Fukui, & Miyazawa, 2015).

2.6.3 Endogenous retroviruses with a role in the host physiology

Besides all of the impacts of the presence of a retrovirus in the host genome listed above, one symbiotic relationship of the host and ERV stands out - endogenous retroviruses domesticated (also called exaptation) to play a role in a normal physiology.

Human endogenous retroviruses were first discovered in human placenta (Kalter et al., 1973). Human ERVW1 expresses its defective env *Syncytin* which was proven to play a role in formation of placenta. *Syncytin* is expressed mainly at placental

syncytiotrophoblasts where it causes the fusion of cells and formation of giant syncytia (Mi et al., 2000). *Syncytin* gene with its fusogenic function was also described in marsupials (Cornelis et al., 2015) and many other mammalian hosts. Multiple ERV variants are present in the reproductive system of Placentalia and have a fusogenic function. The presence in all of the major clades with placenta formation indicates that the presence of fusogenic ERV might be a requirement in order to increase success in reproduction (Imakawa, Nakagawa, & Miyazawa, 2015).

Digging deeper into the Env glycoprotein evolution, an ORF coding for abberant expression *env* with possible role for placentation was discovered in the spiny-rayed fishes. Spiny-rayed fishes are a unique clade, for they evolved placentation. The insertion of the gene is predicted to occur more that 110 MYA, making the detected ORF the oldest gene of retroviral origin (Henzy, Gifford, Kenaley, & Johnson, 2016).

2.7 Expression of endogenous retroviruses

The expression of the integrated provirus is influenced by the environment of the integration site, including the genomic and epigenetic features of the site. No general preference of the integration site of the retroviral DNA was proven, however different groups of retroviruses prefer sites with distinct features (Derse et al., 2007; Elleder, Pavlicek, Paces, & Hejnar, 2002; Mitchell et al., 2004; Narezkina et al., 2004; Schroder et al., 2002; Wu, Li, Crise, & Burgess, 2003).

The transcription of the intgerated retroviruses is often suppressed (retrovirus is silenced). This might occur due to the CpG methylation of retroviral DNA and and is also influenced by various histone modifications (Blazkova et al., 2009; Poleshko et al., 2010). Transcriptional silencing is also one of the key mechanisms leading to the establishment of HIV latent reservoir (Bednarik, Cook, & Pitha, 1990; Blazkova et al., 2009).

2.7.1 Endogenous retroviruses as promoters for mammalian genes

The ERVs present in the host genome are often a source of alternative promoters. However, the presence of ERV does not usually affect the general expression pattern. LTR is rarely preferred as a source of a promoter hence it affects the expression of individual genes minorly, but provides an evolutionary opportunities for modifying the gene expression (Cohen, Lock, & Mager, 2009).

The CRISPR-Cas deletion of ancients ERVs from human genomes in three cell lines led to impaired pathways of IFN signaling. This might be owed to the missing promoters for the genes utilized in the pathway (Chuong, Elde, & Feschotte, 2016).

2.8 Embryonic stem cells and endogenous retroviruses

As mentioned previously, retroviruses need to infect germ line cells in order to endogenize. The analysis of transcription occurring in induced pluripotent murine stem cells revealed that majority of the transcripts originates in LTRs and contains fragments of murine endogenous retroviruses. It was also shown that murine embryonic stem cells and their pluripotency might be related to the presence of endogenous retroviruses and their expression, hence be shaping the early stages of mammalian development (Macfarlan et al., 2012).

This phenomenon was also studied in human stem cells. The embryonic stem cells express HERV-H RNA abundantly. This transcription occurs in reduced amount in induced pluripotent stem cells, and is almost absent in differentiated cells. HERV-H is also regulated in early development and bears binding sites for pluripotency transcription factors. These facts indicate that endogenous retroviruses play a role in early stages of development and cell differentiation (Santoni, Guerra, & Luban, 2012).

3 MATERIALS AND METHODS

3.1 Methods related to the screen of unusual endogenous retroviruses

3.1.1 Endogenous Lentivirus in Malayan colugo

3.1.1.1 Computational screen of vertebrate genomes

The first step of the best bidirectional hit (BBH) strategy was performed by tBLASTn (Johnson et al., 2008) search in vertebrate genome database (including 104 vertebrate genomes available at GenBank) to identify candidate endogenous Lentivirus fragments. In this step the following Pol amino acid sequences (employed for the screen because *pol* is the most conserved retroviral gene) 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 (Johnson et al., 2008) search was set at E-value < 10^{-5}. To filter out non-lentiviral sequences, translated hits were used as a query for backward BLASTp (Johnson et al., 2008) search against database of retroviral Pol sequences belonging to all retroviral genera. Hits aligned with the best bit score to lentiviral sequences in the backward BLAST search were further analyzed.

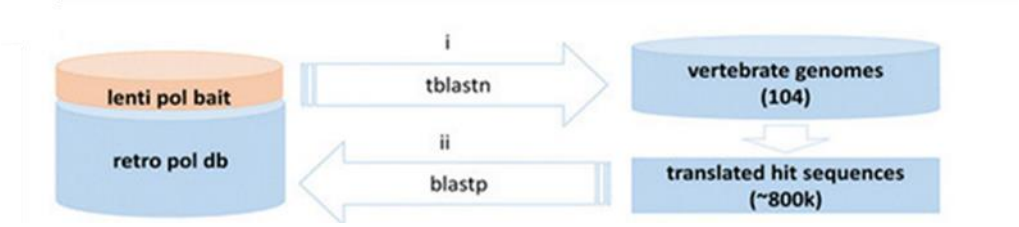

Figure 9: Screening strategy used for screening vertebrate genomes

3.1.1.2 Source of the analyzed samples

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 (designed as GVA3, and GVA5) and *Cynocephalus volans* (CVO).

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 (see Table 4). The *CHRNA1* locus was verified by sequencing of the PCR products in all specimens and comparison with previously described sequences.

3.1.1.2.1 Whole genome amplification (WGA) of DNA samples

Due to the low amounts available, dermopteran DNA samples were amplified by whole-genome amplification for majority of the applications using illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare) according to the manufacturer's protocol.

3.1.1.3 PCR amplification

The ELVgv RT region was amplified using two primer pairs, ELVgvF1/R1 and ELVgvF2/R2, yielding PCR products of around 245 and 215bp. The short junctions between ELVgv 3'end and the host 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-integrations 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 into pGEMTeasy vector (see 3.2.3 DNA cloning and vectors used) and multiple clones were re-sequenced. For the primer sequences see Table 4.

The general PCR conditions for a typical 20 µl reactions were the following: 1x Phusion polymerase buffer, 200 µM d NTPs, 0,5 µM forward and reverse primer (each), 2% DMSO, 0,5 units of Phusion polymerase (New England Biolabs), and 200ng WGA amplified DNA.

The reactions were run according to the following protocol: initial denaturation at 98°C for 30 sec; 33 cycles of 98°C for 8 sec, 54°C for 25 sec, 72°C for 1 min (per 1kbp amplified); the final elongation at 72°C for 10min.

Table 4: Primers used for ELVgv and control genes amplification and sequencing

Primer	Sequence (5'-3')	Description
ELVgvF1	GAATAATGCAGGGCCAGGTA	ELVgv virus PCR
ELVgvR1	GGGTTTCAAATCCCCACTTT	ELVgv virus PCR
ELVgvF2	TCCTAAGATACAAACAGAAAACATTCA	ELVgv virus PCR
ELVgvR2	GCTTTTGATGGACAGCTCCT	ELVgv virus PCR
Colugo1F	GGGGAACTTTGGCGAAGTGTT	control gene (*FES*)
Colugo1R	TCCATGACGATGTAGATGGG	control gene (*FES*)
Colugo2F	GACCATGAAGTCAGACCAGGAG	control gene (*CHRNA*)
Colugo2R	GGAGTATGTGGTCCATCACCAT	control gene (*CHRNA*)
ColugoLTR	CCCAGAACTTTGTGTCTGGTTT	ELVgv LTR for 3' junction
ColugoA1	TGAGGCATCTCTTTCGGTATTC	ELVgv integration A
ColugoA2	CAGCACAGAATACAAGCAGTAGG	ELVgv integration A
ColugoB1	ACAGCAAATCTCACCCTCCA	ELVgv integration B
ColugoB2	GGGTTGTCCATAAAAACAGAACC	ELVgv integration B
ColugoC1	TCTTGATGTTGGCATCAGTTTG	ELVgv integration C
ColugoC2	CCAGCAAGAAGCTGGACATC	ELVgv integration C
ELVseq1	GCAACCAGAAACCAGACACA	Sequencing the 5' 3kb fragment of ELVgv
ELVseq2	CAGCTGGCTAAAACATATATAAGACAT	Sequencing the 5' 3kb fragment of ELVgv
ELVseq3	CAGGGAAGACAGCACACTGG	Sequencing the 5' 3kb fragment of ELVgv
ELVseq4	GGATTTATGCCAGGTAGAAGTAAAAGG	Sequencing the 5' 3kb fragment of ELVgv
ELVseq5	ACCATGATGGCAGAGGCTTT	Sequencing the 5' 3kb fragment of ELVgv
ELVseq6	ATGATCAAGTTAAAATGACATGGAAT	Sequencing the 5' 3kb fragment of ELVgv
ELVseq7	CATCCAGCAGGGTTTTCACA	Sequencing the 5' 3kb fragment of ELVgv
ELVseq8	GCCTTTTACTTCTACCTGGCATAAA	Sequencing the 5' 3kb fragment of ELVgv
ELVseq9	ATTCCATGTCATTTTAACTTGATCATA	Sequencing the 5' 3kb fragment of ELVgv

3.1.1.4 Phylogenetic analyzes

The alignment was generated in MEGA5 software (Tamura et al., 2011) using the MUSCLE algorithm (Edgar, 2004). The ML tree was constructed in MEGA5 software, using the rtREV amino acid substitution matrix (Dimmic, Rest, Mindell, & Goldstein, 2002), 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 (Huelsenbeck & Ronquist, 2001).

The analysis of phylogenetic relationship of ELVgv to other exogenous and endogenous lentiviruses was based on alignment including 2,350 most conserved nucleotides of *gag-pol* from 31 lentiviruses (Gilbert, Maxfield, Goodman, & Feschotte, 2009), together with ELVmpf (Cui & Holmes, 2012a; Han & Worobey, 2012), and ELVgv sequence. The alignment was generated in MEGA5 program (Tamura et al., 2011) using the MUSCLE algorithm (Edgar, 2004). The ML analysis was performed using MEGA5 program under Tamura-Nei model, Nearest-Neighbor-Interchange ML heuristic method and otherwise default parameters. Bootstrap supports were calculated as a percentage out of 1,000 replicates. 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 maximum likelihood (ML) and Bayesian methods were combined.

3.1.2 Endogenous Deltaretrovirus in the genome of Miniopteridae bats

3.1.2.1 In silico sequence analysis

Sequence datasets available at NCBI SRA from Miniopterid species genome or transcriptome (accession numbers PRJNA270665, PRJNA270639 and PRJNA218524) were queried by BLAST 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 mainly used to correct errors in the original MINERVa-containing contig from the *M. natalensis* genome assembly.

3.1.2.2 Source of the analyzed samples and sample processing

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. The bat species were identified with respect to their external morphological traits and the identification was further confirmed by amplification and sequencing of *cytochrome b* or *RAG2* loci.

The DNA was isolated from the ethanol-preserved samples according to phenol-chloroform extraction protocol (See section 3.2.13).

Table 5: List of bat species analyzed

Latin name	**Common name**	**Family**	**Gene locus used for classification**
Miniopterus schreibersii	Schreibers' long-fingered bat	Miniopteridae	*cytB*
Miniopterus fraterculus	Lesser bent-winged bat	Miniopteridae	*cytB*
Tadarida teniotis	European free-tailed bat	Molossidae	*RAG2*
Myotis myotis	Greater mouse-eared bat	Vespertilionidae	*RAG2*
Eptesicus serotinus	Serotine bat	Vespertilionidae	*RAG2*
Hypsugo savii	Savi's pipistrelle	Vespertilionidae	*RAG2*
Plecotus austriacus	Grey long-eared bat	Vespertilionidae	*RAG2*
Pipistrellus pipistrellus	Common Pipistrelle	Vespertilionidae	*RAG2*
Cistugo seabrae	Angolan hairy bat	Vespertilionidae	*RAG2*
Miniopterus natalensis	Natal long-fingered bat	Miniopteridae	*cytB*
Miniopterus africanus	African long-fingered bat	Miniopteridae	*cytB*
Miniopterus minor	Least Long-fingered Bat	Miniopteridae	*cytB*
Epomops dobsonii	Dobson's fruit bat	Epomophorini	*cytB*
Epomophorus gambianus	Gambian Epauletted Fruit Bat	Epomophorini	*cytB*

3.1.2.3 PCR, sequencing, and further analysis of the genomic DNA

The PCR amplifications were performed with a 1:200 mixture of Deep Vent and Taq polymerases and LongAmp Taq buffer (all from New England Biolabs) with the following conditions: 1 cycle of 4 minute 95°C; 3 pre-amplification cycles of 20 sec 95°C, 2 min 52-55°C (according to primer used), and 1,5 min (per 1kb amplified) 65°C; 30

cycles of 20 sec at 95°C, 30 sec annealing at the temperature 4°C higher than in the pre-amplification steps, 1,5 min (per 1kb amplified) 65°C.

The full-length MINERVa provirus was amplified only from 2 samples (genomic DNA from *M.schreibersii* and *M.fraterculus*) using a nested PCR approach with primers DeltaGF6 and DeltaGR4 in the first PCR run; and DeltaGF5 and Delta GR5 in the second round. The product was sequenced using primers DeltaGR4, DeltaGF5, DeltaF3, DeltaR1, DeltaF1.

The 5' end of the provirus was amplified from the genomic DNA from all of the analyzed samples (M. schreibersii, M.fraterculus, M.natalensis, M.africanus, and M.arenarius) using primers DeltaR1+DeltaGF6. The 5' end product was sequenced using primers DeltaR1, DeltaGF6, DeltaF1, DeltaR6. The 3' end of the provirus was amplified using semi-nested PCR from all of the samples with primers GF6+GR4 in the first round and primers F1+GR4 in the second PCR reaction. The 3' end product was sequenced using primers DeltaGR4, DeltaF1, DeltaF8, DeltaF3, DeltaR3, DeltaR6.

The primers R3 and F1 were used for amplification from genomic DNA to check for the possible presence of a more complete *pol* and *env* gene elsewhere in the genome without yielding any longer product than the one predicted from the single deleted provirus described. The control gene *cytB* was amplified from genomic samples using primers cytBMVZ04 and cytBMVZ05 (Smith & Patton, 1991). The same primers were used for sequencing of this PCR product. Due to the fact that *cytB* amplification was not successful from some Chiroptera species (Tadarida, Myotis, Eptesicus, Hypsugo, Plecotus, Pipistrellus, and Cistugo), we amplified an additional control gene *RAG2* using primers RAG2_968R and RAG2_428F (Smith & Patton, 1991). For the M. schreibersii, M. natalensis and M. fuliginosus we designed species-specific cytB primers denoted cytB_natalR and cytB_natalF.

All of the used primers are listed in Table 6. Desired PCR products were sequenced directly after isolation from the agarose gel. We performed further PCR reactions to confirm the presence/absence of MINERVa proviruses in the examined species using primers which amplify the empty integration sites, LTRs, and *gag* regions of the provirus.

Table 6: Primers used in experiments associated with endogenous Deltaretrovirus in Miniopteridae bats

Primer	Sequence (5'-3')	Primer localization
DeltaF1	GACAAGGGTCGAGTCACCTCCTAA	MINERVa *gag*
DeltaF2	AATCTCTCCTTCTGGCCTCTCACA	MINERVa *gag*
DeltaF6	ATTCATGAGGTGCACGTTTAAGCA	5'flanking region of MINERVa provirus
DeltaF8	TATGTTTCCCCATACCTTGCCATCA	MINERVa LTR
DeltaR1	GAGGTCGCAGGGTTATATGGAGGT	MINERVa *gag*
DeltaR4	GGCATCAAAAGGTAAACAGAAGCA	3'flanking region of MINERVa provirus
DeltaR5	CATGGTTCCACTGGTTATCATTTACA	3'flanking region of MINERVa provirus
DeltaR6	CAATCGGCGGGGAGCTTAC	MINERVa LTR
DeltaF5	GGTGCACGTTTAAGCACATACTCG	5'flanking region of MINERVa provirus
CytBnatalL	GTTGCTCCTCAGAAAGATATTTGTCCTC	Miniopterus cytochrome B locus
CytBnatalR	ATGACCTGTGATATGAAAAACCACTGTTG	Miniopterus cytochrome B locus
DeltaF4	GTTGGTTGCTCTCTTGCC TAGTCG	MINERVa LTR
DeltaF10	GGAATACCCGTTTCAGAGAGCAGA	Miniopterus genomic locus 1
DeltaR9	TGATCCCTGAGATGACAGAAGTCG	Miniopterus genomic locus 1
DeltaF9	TTCAGTATTGTGAAAGGGCTCTGC	Miniopterus genomic locus 2
DeltaR8	TCACTCTCTGGCTTTAGAGTCCTTCA	Miniopterus genomic locus 2
DeltaF7	TCATGTAAATGATAACCAGTGGAACC	Miniopterus genomic locus 3
DeltaR7	TGCAATGTGAGTTGTTGAAAGTGAAA	Miniopterus genomic locus 3

3.1.2.3.1 Sequence assembly

The sequence chromatograms obtained were checked and sequences were edited and assembled using SeqMan software (Lasergene 10.0.0 (http://dnastar.com)).

3.1.2.3.2 Sequence annotation and ORF detection

For sequence annotation, the obtained sequence was aligned with other Deltaretrovirus sequences (HTLV-1 and BLV) annotated in NCBI Nucleotide database. The sequences were aligned using MAFFT algorithm included in the MegAlign software from LASERGENE package 10.0.0 (http://dnastar.com). Splice site acceptor and splice site donor sites were determined using the online prediction algorithm (http://www.fruitfly.org/seq_tools/splice.html). The ORFs were predicted using ExPASy translate tool from the ExPASy portal (http://web.expasy.org/translate/).

3.1.2.3.3 Prediction of RNA secondary structures in LTR

Secondary structure prediction of MINERVa putative Rex Response element was performed. Stem loop prediction in the 5'LTR in MINERVa was compared to other deltaretroviruses. The sequences used for secondary structure prediction in mfold

(http://unafold.rna.albany.edu/) were the following: MINERVa consensus sequence, HTLV-1 (GenBank accession number M37299), and BLV (K02120).

3.2 Methods related to the CrERV project

3.2.1 Tissue cultures

3.2.1.1 Cells & media

All cells were cultured in a humidified incubator at 37°C and 5% CO_2.

3.2.1.1.1 Human rhabdomyosarcoma cell line A-673

Human rhabdomyosarcoma cell line A-673 (ATCC product number CRL-1598) was grown in Dulbecco's modified Eagle's medium (DMEM) (Sigma–Aldrich, St. Louis, MO) supplemented with 10% fetal calf serum, penicillin (100U/ml) and streptomycin (100 µg /ml).

3.2.1.1.2 Human embryonal kidney 293T cells and their derivates

Human embryonal kidney 293T cells (HEK293T) were grown in in DMEM, with serum supplements (4% fetal calf and 4% calf serum or 10% fetal calf), penicillin (100U/ml) and streptomycin (100 µg /ml). HEK293T cells expressing *gag* and *pol* genes derived from MLV virus were grown as unmodified HEK293T cells, but on tissue culture dishes covered with Poly-L-lysine (1mg/ml in H_2O). HEK-293T cells producing PERV 14/220 (Bartosch et al., 2004) were grown as regular HEK293T cells and used as a source of PERV particles. HEK293T cells stably transfected with molecular clone of CrERV were grown as regular HEK293T cells and denoted as Cr5. HEK293T cells producing molecular clone of CrERV and a GFP-encoding vector with two LTRs (denoted as C10 line) were used as a source of CrERV marked with GFP. HEK293T cells stably expressing a viral (MLV-based) construct with amphotropic envelope were maintained in the same medium as regular HEK293T cells.

3.2.1.1.3 Primary cells of Cervidae species

Primary mule deer (*Odocoileus hemionus*) kidney cells (OHK, ATCC product number CRL-6193) were grown in DMEM supplemented with 10% fetal calf serum, penicillin (100U/ml) and streptomycin (100 µg /ml).

Primary red deer (*Cervus elaphus*) and fallow deer (*Dama dama*) cells were grown in DMEM supplemented with 10% fetal calf serum, penicillin (100U/ml) and streptomycin (100 µg /ml).

3.2.1.2 Cryopreservation of the cells

For cryopreservation, the cells were harvested and spun at 200 g for 5 min at 4°C. The cellular pellet was then resuspended in the cultivation medium containing 10% DMSO. The mixture was slowly cooled in the tubes in the freezing box with isopropanol (Nalgene) at -80°C for 24 hours and after that the tubes were stored in liquid nitrogen.

3.2.1.3 Co-cultivation of human and primary deer cells

The co-cultivation experiment was set by mixing equal amount of deer CRL-6193 and human A-673 cells. The cells were kept in the co-culture for several weeks. Every week, fresh cells from both species were added to the coculture at a 1:1 ratio. At indicated time points, samples of the culture medium were harvested for the RT assay. The samples were centrifuged at 3,000 RPM for 5 min and filtered by a 0.22 µm syringe filter in order to remove cellular debris. The samples were frozen at -80°C before further analysis.

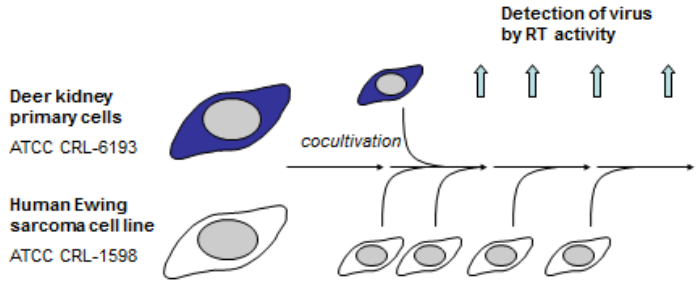

Figure 10: Schematic of the cocultivation experiment

3.2.1.4 Ultracentrifugation methods -Iodixanol gradient

Iodixanol (OptiPrep™) was purchased from Axis Shield (Dundee). Thirty milliliters of cell-free supernatants from virus-producing cells were centrifuged to remove cell debris (3,000 RPM for 5 minutes at 4°C). Subsequently the supernatants were centrifuged through a 20% iodixanol cushion in a SW28 rotor (Beckman Coulter) for 2h at 23,000 RPM. The centrifuged pellet was resuspended in 1 ml of ice-cold PBS. Two-

milliliter layers containing 50%, 40%, 30%, 20%, and 10% Iodixanol were layered in tubes for the SW41Ti rotor (Beckman Coulter) and the resuspended pellet in PBS was applied on top of the prepared gradient. The gradient was centrifuged for 17 h at 35,000 RPM at 4°C. Twelve gradient fractions were collected from the top and their density was determined by refractometry. Aliquots from each fraction were used for the PERT assay.

3.2.1.5 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).

3.2.2 PCR methods

3.2.2.1 Conventional PCR

Each PCR reaction mixture had a total volume of 20 µl, containing 1.5 µl of the DNA (up to 0.5 µg) solution and 300 nM (each) the forward and reverse primers, and 160 µM of each dNTP. The PCR amplifications were performed with One Taq polymerase and its according buffer (all from New England Biolabs) with the following conditions: 1 cycle of 4 min 95°C; and 25 cycles of 15 sec 94°C, 25 sec 55-60°C (according to the primer used), 1 min (per 1kb amplified) at 68°C; final elongation for 10 min at 68°C.

For determination of an integration pattern of CrERV, ExTaq polymerase (Takara) was used with the following PCR program: 1 cycle of 95°C for 2 min, and 31 cycles of 94°C for 20 sec, 55°C for 30 sec, 68°C for 2 min, and final elongation for 10 min at 68°C. The primers used were Prp1 and Cervus2 (see Table 7).

Table 7: Primers used for CrERV amplification, insertion pattern determination and sequencing.

Primer	Sequence (5'-3')
MDfor5	GACAATTCCGGCAAACAATAA
MDRe6R	CGAGGACGGAGGTTTTGGAGC
MDfor5	GACAATTCCGGCAAACAATAA
QMDgagR1	GCAAGAGGCATCCTGAAAGA
MDfor4	CCCAGAGTTACCGTCATCCA
MD1810R	CAACTGTTGGCAATCATCCC
MDpol704F	GACTGCAGAGGAAAGGGAAGAACG
sMDpol1R	GATGGACTTGGACAACCGATA
sMDpol2F	CAGTGGCTGCAATGATGC
sMDpol1F	CTGTCCGAAATCCCCTTTCT
MDRint4R	ACAAAGGCAGGTCCGTTATCAGAG
QMDgagR2	TGTATAAGGCCCGATTTTCG
QMDgagF1	CCAGGTCCCTTATATCGTGGT
MDPpol2391R	ACCAAGTGTACTCAGGCAGCAGA
Prp1	GCAACCCATTCCAGTATTCTT
Cervus2	TGGTAGAGAGAACGCAATGG

3.2.2.2 Real-time quantitative PCR

MESA GREEN qPCR mastermix (Eurogentec) 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^{TM} Real-Time System (Bio-Rad) with a two-step protocol (1 cycle of 5 min at 95°C and then 44 cycles consisting of 15 sec at 95°C and 60 sec 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 human genomic locus in porphobilinogen deaminase gene (Konig et al., 2008).

Table 8: Primers used for qPCR to quantify the amount of CrERV env

Primer	Sequence (5'-3')
QMDenvF1	TGACCCCATGTTTGAATGTG
QMDenvF2	CAAACCAAGGAGCTGTCCTC
QMDenvR1	GAGGACAGCTCCTTGGTTTG
QMDenvR2	CCCACCTTGCTGAAGAAAAA

3.2.2.3 Digital droplet PCR

For highly accurate absolute quantification of viral DNA, droplet digital PCR (ddPCR) system QX200 (Bio-Rad) 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 sec at 95°C and 40 sec 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.

*None of the data obtained by ddPCR are presented in this thesis, but can be found in the publications connected to this thesis.

3.2.2.4 Product enhanced reverse transcriptase (PERT) assay

The PERT assay was adapted from previously published protocols (Lovatt et al., 1999; Pizzato et al., 2009; Sears & 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 minutes. Then, two master-mixes were prepared, with the following amounts per one reaction: Mix 1 contained 20 ng of the template RNA of MS2 phage (Roche), 0.5 µl of the reverse primer MS2b (5'- GCCTTAGCAGTGCCCTGTCT) and 10.1 µl water. Mix 2 contained 3.6 µl of 5x ProtoscriptII buffer, 2 µl of 100 mM DTT, 0.8 µl of 10 mM dNTP2, and 6.4 µl water. Mix 1 was incubated at 65°C for 5 minutes and slowly cooled down to allow primer annealing. Next, the mixes were pooled and aliqoted by 18 µl. To each aliquot, 2 µl of the lysates were added and incubated at 37°C for 30-60 minutes (reverse transcription step), then inactivated at 70°C for 10 minutes. The newly generated MS2 cDNA was quantified by real-time PCR assay with forward MS2a (5'-AACATGCTCGAGGGCCTTA) and reverse MS2b primers and fluorescent probe (FAM-TGGGATGCTCCTACATG-TAMRA). Each reaction contained 1.5 µl of the cDNA sample, 1xqPCR master mix (Eurogentec), 7.5 pmol of each primer and 3.75 pmol of probe in a total volume of 15 µl. The samples were run on a Bio-Rad CFX96^{TM} Real-Time instrument with a three-step protocol: 1 cycle of 10 min at 95°C and then 45 cycles consisting of 15 sec at 95°C, 20 sec at 60°C and 20 sec at 72°C. Cycles of quantification (Cq) values were

generated by the CFX Manager software. With each run, one calibrator sample (MLV virions pseudotyped with VSV-G) was assayed and all values were expressed as relative values compared to the calibrator.

3.2.3 DNA cloning and vectors used

3.2.3.1 Preparation of DNA inserts

3.2.3.1.1 Plasmids with env derived from CrERV

CrERV *env* was amplified from OHK cell line genomic DNA using primers CrENVe1 (5'-cttaagcttccaccATGGAAGGCGAATGCTCATC) with ctt overhang (green), HindIII restriction site (underlined), Kozak sequence (blue) and proviral *env* (start codon red and other coding sequence yellow) and downstream primer CrENVe2 acgttgaattcTTATGGGGAGGAATCTTCCTCT with overhang (green), *Eco*RI restriction site (underlined) and stop codon (red).

CrENVFLAG was amplified from subcloned CrERVenv vector with primer CrEnvApa (5'-CTCCTGGGCCCACTTTTACT) with *Apa*I restriction site (underlined) and primer CrEnvFLAG, (5'-AAgggcccTCACTTGTCATCGTCGTCCTTGTAGTCTCCTGGGGAGGAATCTTCCTCT) amplified from the genomic DNA sample AGAGGAAGATTCCTCCCCAGGA*GACTACAAGGACGACGATGACAAG*TGAgggcccTT where yellow is the *env* sequence, glycine is gray, FLAGtag is green (stop codon red) and ApaI restriction site is underlined

CrERVnT was amplified from subcloned CrERVenv vector using CrEnvApa primer and primer CrENVnT (5'-AAGGGCCCTCAATTTAAAATACAAGGCCCAATTGTAA) targeting the sequence CTCCTGGGCCCACTTTTACTGTTAATACTAATACTTACAATTGGGCCTTGTATTTTAAATTGAgggcccTT introducing an early stop codon (red).

3.2.3.1.2 Plasmids with env derived from amphotropic MLV

Amphotropic *env* was amplified from commercial plasmid PCL Ampho (Addgene) using primer AmphoENV1 (5'-aatgGATCcACCATGGCGCGTTCAACGCT) with *Kpn*I restriction site (underlined) and Kozak sequence (blue) and primer AmphoENV2 (5'-gctctaGATCATGGCTCGTACTCTATGG) with *Xba*I restriction site (underlined).

AmphoenvFLAG was amplified from the subloned Amphotropic env vector using primer AmphoCla1 (5'-TGCATTCTCAATCGATTAGTCC) with *Cla*I restriction site (underlined) and primer AmphoenvFLAG (5'-ATTCTAGATCACTTGTCATCGTCGTCCTTGTAGTCTCCTGGCTCGTACTCTATGG) generating a sense strand sequence CCATAGAGTACGAGCCAGGA*GACTACAAGGACGACGATGACAAG*TGATCTAGAAT with glycine (gray), FLAGtag (green, stop codon red) and XbaI restriction site (underlined).

3.2.3.2 Plasmid vectors

For DNA cloning procedures, the plasmid pGEM-T Easy (Promega) was used. For constructing expression vectors, expression plasmid pcDNA3 (Addgene) was used.

3.2.3.3 Ligation

All DNA inserts were first subcloned into the pGEM-T Easy vector according to the manufacturer's protocol using molecular ratio insert to vector 3:1; transformed to E. Coli (see chapter bacterial transformation), the colonies were screened by PCR and the products were sequenced. The constructs were then digested and re-cloned to expression vector pcDNA3. The constructs were again transformed to bacteria, colony screened and sequenced. Prior to ligation of the CrERVenvFLAG construct with pcDNA3 plasmid, the plasmid was dephosphorylated using rAPiD alkaline Phosphatase (Sigma Aldrich) according to manufacturer's protocol.

3.2.3.4 Bacterial transformation

The plasmids with cloned inserts were transformed to E.Coli strain XL1 blue (if restriction enzymes were sensitive to bacterial dam methylation, SCS110 strain was used). Aliquots of the ligation mixture were incubated with bacteria on ice for 40 minutes. After that heat shock at 42°C was peformed for 1 minute. The mixture was immediately added to 0.5 ml of the LB media and recovered for 40-60 minutes at 37°C. The recovered bacterial culture was seeded to LB-agar plates with ampicilin (1 µg/ml, Sigma), IPTG (3mg/ml, Thermo Scientific), and S-gal or X-gal (both 30 mg/ml, Thermo Scientific), to enable blue/black-white colony screening. The plates were incubated overnight at 37°C.

3.2.3.5 Colony screen PCR

The white colonies were transferred to 20 µl of LB media with ampicillin. The 2 x OneTaq master mix (New England Biolabs) was used to mix reactions of 15 µl containing 1.5 µl of bacterial mixture. Primers designed to anneal to the end of bacterial vectors were used: M13F and M13R for pGEMTeasy vector and pcDNA3F and pcDNA3R for pcDNA3 vector.

3.2.4 Other used plasmid vectors

- *pVSV-G* is a 6.5 kbp long plasmid bearing an envelope protein of VSV-G (vesicular stomatitis virus G) with cytomegalovirus (CMV) promoter.
- *pLG* is a 5.658 kbp long plasmid bearing two MMLV and MMSV LTRs and EGFP with Ampicilin resistance and enhanced packaging singnal.
- *pBS-CMV-gagpol* is a 9.333 kbp long plasmid bearing an MLV *gag-pol* available from Addgene (catalogue number 35614).

3.2.5 Plasmid DNA isolation

For the isolation of plasmids for sequencing or re-cloning, QIAprep Spin Miniprep Kit (Qiagen) was used according to manufacturer's spin protocol.

The plasmids used for transfection were grown in bacteria in LB medium and isolated using GenElute HP Plasmid DNA Midiprep Kit (Sigma Aldrich) according to manufacturer's protocol. The isolated plasmids were stored at 4°C.

3.2.6 Storage of the transformed bacteria

The bacteria bearing desired plasmids grown in LB medium overnight were mixed with 10% glycerol in 7:3 volume ratios and subsequently stored at -80°C.

3.2.7 Transfection

For obtaining the retroviral constructs in sufficient amounts, calcium phosphate transfection was used. Cells were seeded to reach approximately 60% confluence on a 100 mm plate. Up to 30 µg of DNA was dissolved in 1080 µl of water, mixed with 135 µl 2M $CaCl_2$ and subsequently dripped stepwise to the mixture of 1120 µl 2xHBS and 22 µl 100xPO_4 while being lightly shaken to let the mixture form a precipitate; then the mixture was added to the cells (with medium changed at least one hour prior to the

procedure) in a dropwise manner. After 5 h incubation, cells were washed with warm (37°C) 15% glycerol in 1xHBS for 30 sec, washed with PBS and supplied with a fresh medium. The medium with the product of transfection was collected on day two and three after transfection. It was spun at 200 g for 10 min at 4°C to get rid of cell debris and immediately used or stored at -80°C.

3.2.8 Infection

For infections, medium from virus-producing cells was harvested, spun for 10 min at 4°C at 200 g to remove cell debris and either used to infect the cells directly or was stored at -80°C. Medium harvested from cells expressing a reasonably high viral titer was filtered through 0.45 μm filter.

If a virus was concentrated prior to use, it was spun at 23,000 RPM for 2 h in a 0-micron environment at 4°C in a SW28 rotor (Beckman Coulter).

Virus vectors with an envelope not stable in ultracentrifugation procedures (e.g. Env of amphotropic viruses) were concentrated using RetroConcentin (System Biosciences) according to manufacturers' protocol concentrating the viral stock 20x.

In case RetroNectin (Takara) was used to enhance the viral infectivity, the culture dishes were coated with RetroNectin according to manufacturer's spin protocol. RetroNectin reagent is a recombinant human fibronectin fragment that contains three functional domains. RetroNectin enhances retrovirus-mediated gene transduction by aiding the co-localization of target cells and viral particles. Specifically, virus particles bind RetroNectin via interaction with the H-domain, and target cells bind mainly through the interaction of cell surface integrin receptors VLA-5 and/or VLA-4 with the fibronectin C-domain and CS-1 sites, respectively. By facilitating close physical proximity, the RetroNectin reagent can enhance viral-mediated gene transfer to target cells expressing integrin receptors VLA-4 and/or VLA-5.

In case polybrene was used to enhance retroviral infection, it was added directly to the viral stock to obtain a medium with a polybrene concentration of 4 μg/ml. Polybrene is presumed to enhance the retrovirus uptake in the cell in the assays for retrovirus gene transfer (Davis, Morgan, & Yarmush, 2002).

3.2.9 Retrovirus envelope pseudotypes

Retroviral pseudotypes were constructed by co-transfecting plasmids into HEK293T cells by calcium phosphate transfection. The constructs with concentration of individual plasmids are listed in Table 9.

Table 9: Amounts of plasmid constructs used in the production retroviral pseudotypes

Pseudotyope designation	Envelope construct	Gag-pol MLV plasmid	GFP reporter plasmid (pLG)
CrERVenv	5 µg	10.75 µg	12.5 µg
CrERVenvFLAG	5 µg	10.75 µg	12.5 µg
CrERVenvnT	5 µg	10.75 µg	12.5 µg
Ampho	5 µg	10.75 µg	12.5 µg
AmphoenvFLAG	5 µg	10.75 µg	12.5 µg
Gag-pol	0 µg	25.75 µg	12.5 µg

The medium from the transfected cells was harvested two or three (or both) days after transfection and used for infection of cells.

3.2.9.1 Western blot

The cell lysates harvested from tissue culture plates were applied to a 8% (upper and 12% (lower) SDS polyacrylamide gel and transferred to PVDF membrane, which was subsequently blocked in 5% reduced milk. Membrane was subsequently incubated with an anti-FLAG antibody and a secondary anti-rabbit antibody with covalently bound horseradish peroxidase (Cell Signaling Technology). The luminescent reaction was performed using LumiGLO solution (Cell Signaling) and exposed on an X-ray film (Agfa). Coomasie brilliant blue staining of gels was performed as loading control.

3.2.10 Retrovirus marker rescue assay

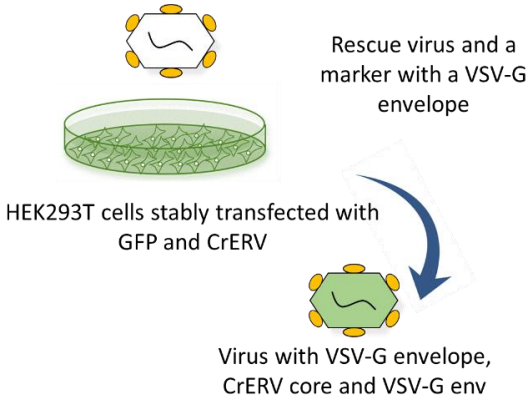

Figure 11: Strategy used for the marker rescue assay. The cells bearing CrERV were transfected with an expression plasmid with VSV-G envelope. HEK293T cells stably transfected with GFP and infected with CrERV were transfected with a plasmid with a VSV-G envelope to produce a virus with CrERV core and Env, GFP marker, and VSV-G.

To perform the marker rescue, the HEK293T cell line bearing a molecular clone of CrERV and GFP was transfected with pVSV-G plasmid. The standard protocol for calcium phosphate transfection was used. The idea behind the strategy of marker rescue is depicted in the Figure 11.

2ml of fresh C10 cells (cells used to generate the rescued GFP-marked virus) medium was used for infection or 1ml of cells medium of CrERV positive cells transfected with VSV envelope. Three days after infection, cells were fixed and stained with DAPI (4',6-Diamidine-2'-phenylindole dihydrochloride) to mark the nuclei.

3.2.11 Flow cytometry (FACS)

Infected cells were harvested and centrifuged (200 x g, 5 min, 22°C), the cellular pelet was resuspended in Hoechst 33258 in PBS (1 µg/ml). Uninfected cells were treated the same way and used as a negative control to distinguish GFP+ and GFP- cells. The number of Hoechst-negative (live) and GFP-positive cells was measured using BD LSRII (BD Biosciences). The data were analyzed using FlowJo software (http://www.flowjo.com).

For fixation, the cells were resuspended in 1% paraformaldehyde (PFM) in PBS after harvesting and centrifugation. Prior to measurement, they were strained through 50 µm strainer (Cell Trics, Sysmex) in order to remove the cell clumps.

3.2.12 Cell fixation for DAPI staining

The cells were seeded on a cover slip in a 35 mm Petri dish and washed three times after 24 hours with PBS (heated to 37°C). The cell were fixed with 2 ml of 3% PFM in PBS (37°C) and incubated for 20 min at a room temperature. The slips were washed with PBS three times (with 15 min between washes), then washed in a demineralized water. A drop of Mowiol+DAPI was placed at the microscope slides. The cover slips were placed on the miscroscope slides so that no air bubbles would be introduced between the slides and slips and surplus fluid was removed.

3.2.12.1 Preparation of 3% paraformaldehyde

3 grams of PFM were dissolved in 70 ml PBS and mildly heated. The pH was adjusted to 6.9 using KOH. The PBS was added up to 100 ml, the solution was filtered and stored at -20°C in aliquots.

3.2.13 Genomic DNA isolation by phenol-chloroform extraction

The cellular pellet was resuspended in 800 µl of lysis buffer (containing water; 0.5M EDTA-pH 8.0; and 10% SDS in the ratio 4:5:1). 40 µl of proteinase K (20 mg/ml) was added and the mixture was incubated at 55°C overnight in a rolling tube. The next morning, 2.5 µl of RNase A (100 mg/ml) was added and the mixture was incubated in the rolling tube at 37°C for 1 h. 800 µl of phenol:chloroform (1:1, pH 7.9) was added and thoroughly but not vigorously mixed. The mixture was subsequently centrifuged at room temperature at 16,000 g for 3 min. Aqueous phase-containing DNA was collected and mixed with an equal amount of cold (-20°C) 96% ethanol. The mixture was mixed and centrifuged (16,000 g, 15 min, 12°C) and the supernatant was removed. The pellet was washed with cold (-20°C) 80% ethanol and centifuged (16000 g, 15 min, 12°C). After removal of the supernatant, the pellet was air dried at 37°C and dissolved in 50 µl $T_{10}E_{0.1}$ (10 mM TRIS, 0.1 mM EDTA) buffer.The samples were stored at 4°C.

3.2.14 Alignments & construction of the phylogeny trees

Alignments of sequences were generated using the MUSCLE algorithm included in the MEGA6 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013) software package. The Neighbor-joining trees were constructed using algorithms in MEGA6 and visualized

in FigTree. The ProtTest analyzes trees were constructed using online ProtTest2.4 server (Abascal, Zardoya, & Posada, 2005).

3.2.15 Assembly of the restriction factors sequences

Sequence datasets available from the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) from mule deer (BioSample PRJNA79789) were queried by BLAST (Johnson et al., 2008) using various ungulate restriction factor sequences as baits. The restriction factors with at least three hits to the mule deer SRA sequences were further processed. Sequencing data from lymph nodes of mule deer, sika deer, and European roe deer were used to assemble selected restriction factors using CLC genomics Workbench 9.5 (http://www.clcbio.com) or DNASTAR Lasergene 10.0.0 (http://dnastar.com). The assembled sequences and sequenced mined from the NCBI Nucleotide database were used for alignment construction, using MEGA6 software (Tamura et al., 2013) and further analyzed using DataMonkey algorithms (Delport, Poon, Frost, & Kosakovsky Pond, 2010) to detect the signatures of positive selection using the default parameters.

3.3 Methods related to the CHOK1 cells project

3.3.1 Tissue cultures

All cells were cultured in a humidified incubator at 37°C and 5% CO_2.

3.3.1.1 The source of amphotropic virus

S. Indik (University of Veterinary Medicine Vienna, Austria) provided HEK293T cells stably expressing a MLV-based retroviral construct with amphotropic envelope. The cells were maintained in the same medium as regular HEK293T cells. The virus-containing cell culture medium was collected and spun to remove the cell debris (200 g, 5 min, 4°C).

3.3.1.2 Cultivation of CHOK1 cells

The CHOK1 cells were maintained in NP medium (composition: 100 ml water; 104 ml D-MEM/F12 2x; 4.5 ml 5% $NaHCO_3$; 10.5 ml fetal calf serum; 10.5 ml calf serum; penicillin (100U/ml) and streptomycin (100 µg /ml); glutamine).

3.3.1.2.1 Tunicamycin treatment

Cells were seeded at a 10% confluency. After four hours, tunicamycin was added to a final concentration of 0.4 μg per ml of medium. The infection was performed 19 h following the tunicamycin treatment.

3.3.1.2.2 Harvesting of the conditioned medium

In order to prepare sufficient amount of conditioned serum free medium we cultivated the CHOK1 cells grown to approximately 80% confluence in serum free NP medium for 24 hours.

3.3.2 Infection

RetroNectin (Recombinant human fibronectin fragment, Takara) coating of the dishes was shown not to increase infectivity; hence it was not used in the experiment. Polybrene was used to enhance retroviral infection; it was added directly to the viral stock to obtain a medium with a polybrene concentration of 4 μg/ml.

To prove the inhibitory effect of the conditioned CHOK1 cells medium (or its fractions obtained by gel filtration), the conditioned medium (or the fractions) and prepared viral stock were added to the cells simultaneously in 1:1 volume ratio. As a control, fresh medium was added to the cells simultaneously in 1:1 volume ratio with prepared viral stock.

3.3.3 Evaluation of the infection outcome

To evaluate infection (and its inhibition) outcome, cells were either analyzed by flow cytometry or the number of foci of GFP-positive cells was manually counted under fluorescent microscope (both approaches discussed in the methods related to CrERV experiments).

3.3.4 Medium concentration, fractionation, and mass spectrometry

The Laboratory of Structural Biology (BIOCEV) performed medium concentration and fractionation. The mass spectrometry was performed by the core facility at BIOCEV.

500 ml of CHOK1 conditioned media was concentrated to 12 ml using Amicon ultrafiltration device (MW cutoff = 10 kDa). 4 ml frozen aliquots of the concentrated

medium was stored at -20°C. 2x4 ml was separated on Superdex HR 75 16/60; fractions were collected by 3 ml.

3.3.5 Evaluation of the mass spectrometry results

The protein sequences obtained were used in BLAST searches against NCBI, TrEMBL, and SwissProt database and a personalized amino acid sequence database. The personalized database was created by predictions of endogenous retrovirus env sequences present in the CHOK1 genome. The best BLAST hits against hamster genome were obtained using the following bait sequences:

Amphotropic_murine_leukemia_virus_strain_1313_AF411814.1,
Amphotropic_Murine_leukemia_virus_4070A_M33469.1,
Ampho_Cricetulus_griseus_env_U09104.1,
Ampho_FIV_EF455613.1,
Ecotropiv_MLV_env_KJ668270.1,
Xenotropic_MLV_env_M59793.1,
Polytropic_MLV_env_KJ668271.1,
KoRV_NC_021704.1, GaLV_NC_001885.2,
GaLV_SEATO_AF055060

4 RESULTS AND DISCUSSION

4.1 Screens for the unusual endogenous retroviruses in mammalian genomes

4.1.1 Endogenous Lentivirus in Malayan colugo (*Galeopterus variegatus*)

We performed a computational screen of 104 mammalian genomes available at the time, aimed at detecting unusual cases of endogenous retroviruses, including endogenous lentiviruses (see Figure 12) and deltaretroviruses. We detected a novel endogenous Lentivirus in the Malayan colugo genome, denoting it ELVgv. We were able to assemble three endogenous copies of ELVgv. Construction of the phylogenetic trees confirmed that the provirus is sequentially most related to lentiviruses. However, ELVgv is evolutionary old and does not clearly cluster with any Lentivirus group, forming a separate group within lentiviruses.

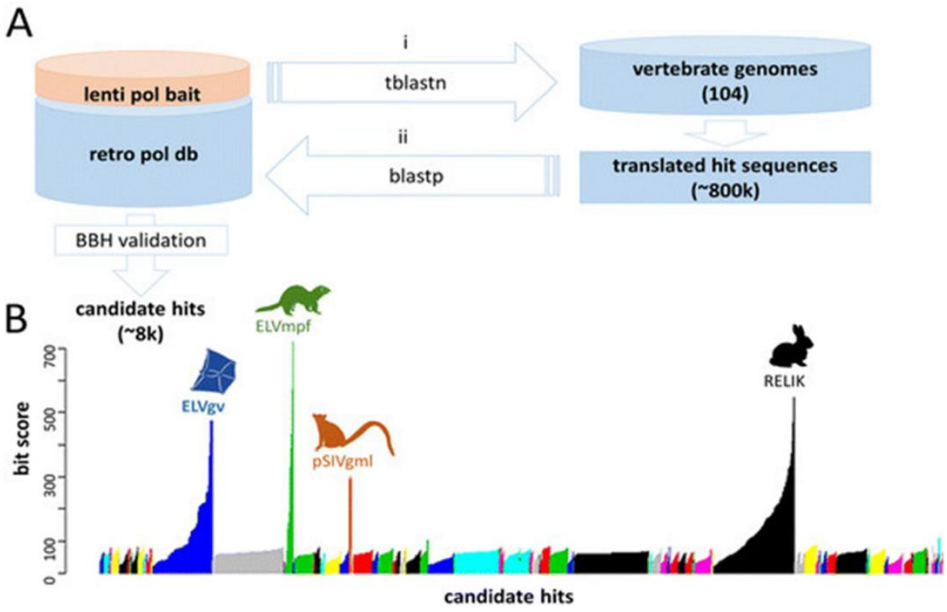

Figure 12: A. Schematic of the computational pipeline used for screening the vertebrate genomes. B. The outcome of the analysis. Every blast hit is denoted by 1 vertical line. The height of the line depicts the value of the bit score of the hit. The individual species are distinguished by various colors. Figure from (Hron et al., 2014).

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). All three detected proviruses in the genome are displayed in the Figure 14. The majority consensus sequence of the proviruses was used to construct the phylogenetic trees depicted in the Figure 13.

To estimate the phylogenetic relationship of ELVgv to other lentiviruses, we have constructed phylogenetic trees using alignment of the amino acid sequences of reverse transcriptase. Reverse transcriptase is a highly conserved region of *pol* with sequences available 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), see Figure 13 for reference. 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).

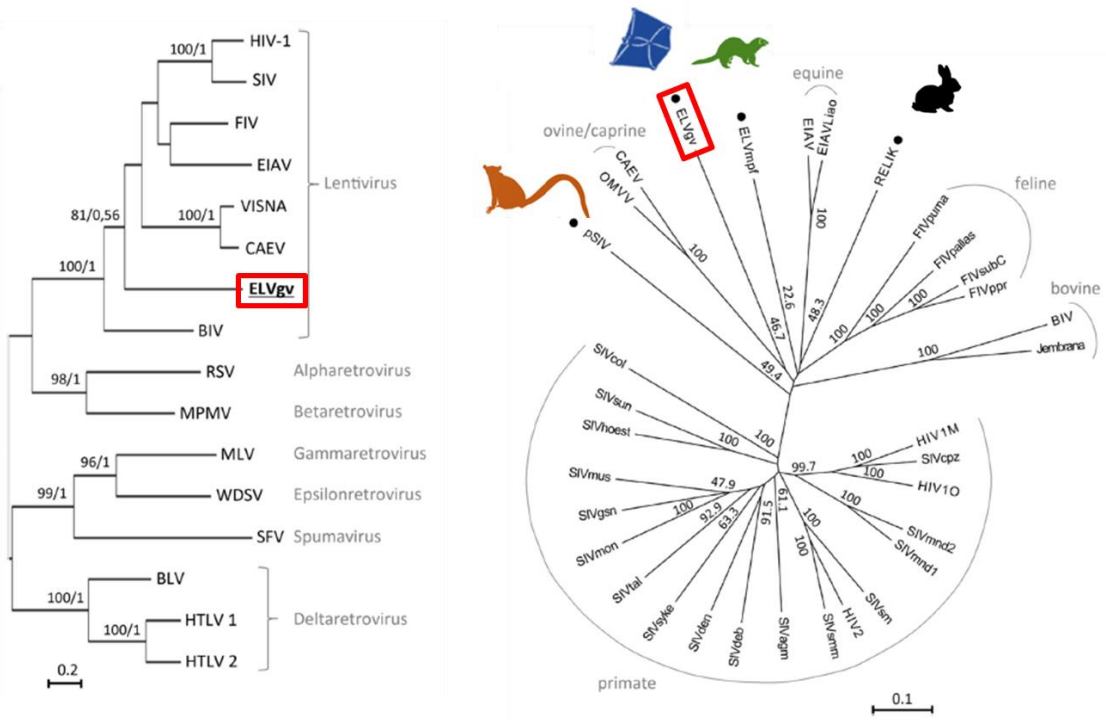

Figure 13: Phylogenetic analysis of ELVgv. The sequence of ELVgv clusters with good support with other members of Lentivirus genus. Bootstrap supports as well as posterior probability of Bayesian tree is displayed at the nodes (left). The situation within the lentiviral genus is depicted as well (right). The endogenous lentiviruses are indicated by a dot. None of the endogenous forms of lentiviruses clusters in a group with recent lentiviruses. The numbers represent the bootstrap values.

To analyze the relationship of ELVgv to other lentiviruses in more detail, we have used the dataset of conserved regions of lentiviral *gag* and *pol* (Gilbert et al., 2009), adding the recently described ELVmpf to the analysis (Cui & Holmes, 2012a; Han & Worobey, 2012). The ML phylogeny analyzes indicate that ELVgv forms a deep branch in the Lentivirus tree, indicating its ancient origin and distinct position in the lentiviral genus. ML tree differed slightly from the phylogeny obtained by Bayesian analysis. While in the ML analysis ELVgv clustered with weak bootstrap support (46.7) together with the ovine/caprine Lentivirus subgroup, it formed an isolated deep branch in the Bayesian tree. Separate analysis of the *gag* and *pol* genes excluded any evident recombination event. Re-running the analysis with the three individual provirus sequences instead of the reconstructed ELVgv consensus sequence also did not influence the results (shown in the supplementary material of (Hron et al., 2014)). Based on these facts, the precise relationship of ELVgv to primate and nonprimate Lentivirus groups could not be determined.

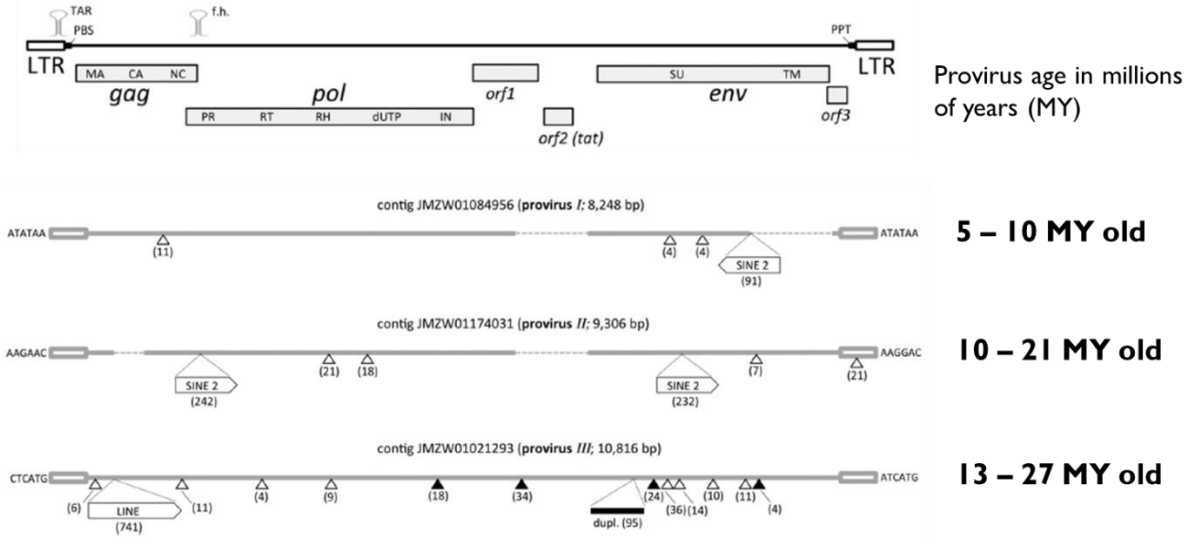

Figure 14: The assembly and structure of three ELVgv proviruses in the colugo genome; age estimation of individual ELVgv integrations based on LTR aging is depicted on right. Figure from (Hron et al., 2014).

Dermoptera order contains only two extant species - *Galeopterus variegatus* and *Cynocephalus volans*. Dermoptera is considered the basal branch to primates. We obtained dermopteran genomic DNA samples in very low amounts. The samples were amplified by whole genome amplification and their correct identification was validated

by sequencing *CHRNA1* control gene and comparing the sequences to the ones published in GenBank (alignment depicted in Figure 15). The identity of all samples was confirmed in this way.

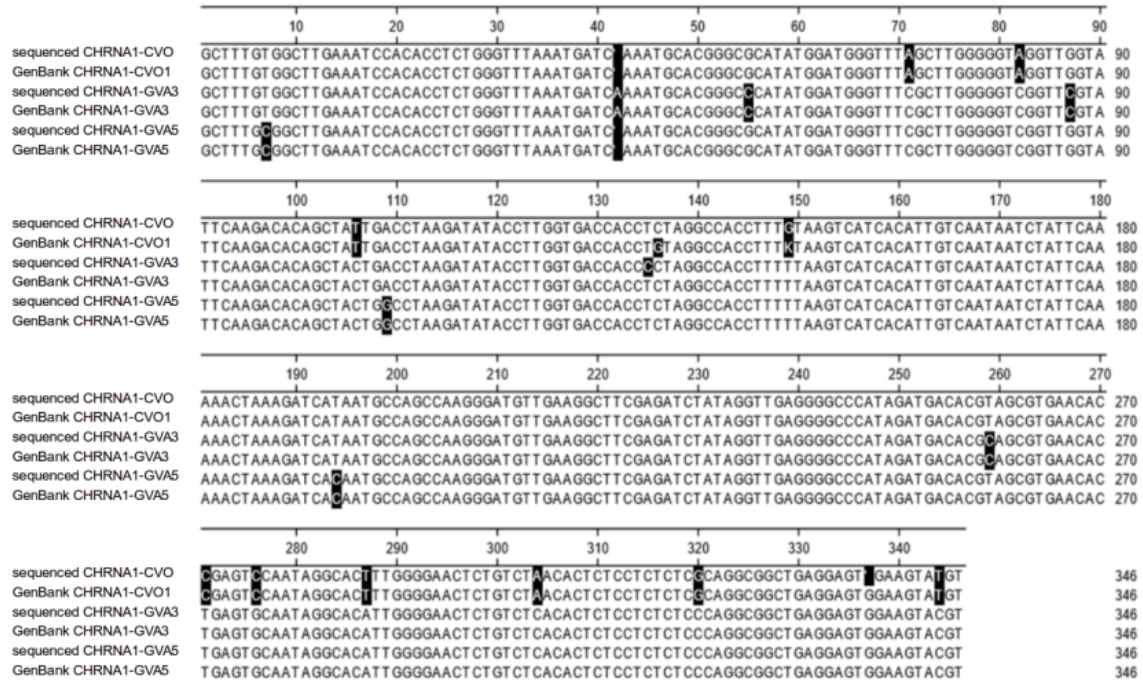

Figure 15: Sequence alignment of CHRNA1 control locus from dermopteran specimens analyzed . CHRNA1 locus was PCR-amplified and sequenced from dermopteran genomic DNA to confirm the identity of specimens analyzed. Sequences with the following GenBank accession numbers were used for comparison: CVO1 CHRNA1 (FJ151285), GVA5 CHRNA1 (FJ151283) and GVA3 CHRNA1 (FJ151281).The polymorphic sites used for comparison are highlighted in black.

The presence of ELVgv in *Galeopterus* and *Cynocephalus* genome was confirmed by PCR detection of RT region from whole genome amplified (WGA) samples. The presence of ELVgv in both of the species confirms the prediction, that ELVgv integrated in the genome of ancestor of the currently living dermopteran species more than 20 MYA (see Figure 16).

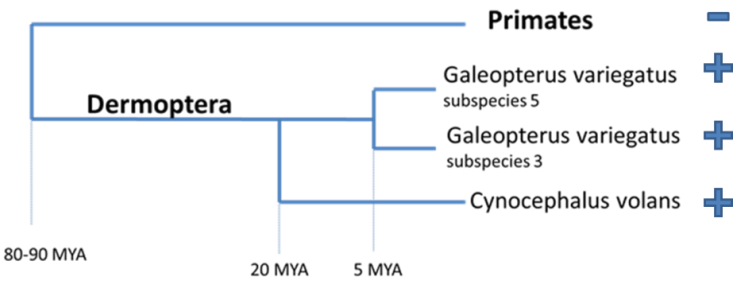

Figure 16: Presence/absence of ELVgv among species. The numbers associated with the nodes depict the time of the divergency of the species according to TimeTree (http://www.timetree.org/).

Further studies of the ELVgv confirmed the estimated age of the provirus and marginally dealt with the host-virus relationship. We were able to detect and sequence partial provirus from the other extant species from the Dermoptera clade. We partially sequenced the three endogenous copies of the provirus in three specimens. We looked for empty integration site from all of the proviruses, but detected none (see Figure 17). This fact indicates that the provirus insertion is homozygous and in the same genomic loci in all animals tested.

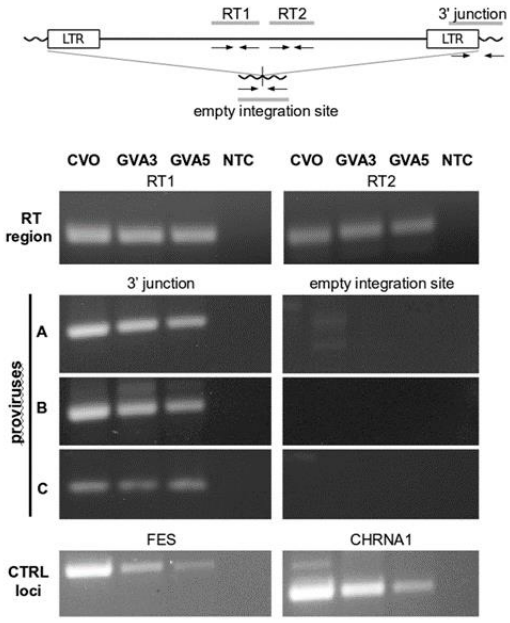

Figure 17: Detection of ELVgv integrations in the dermopteran genomic samples. Upper*: schematics of ELVgv provirus with primer positions.* Lower*: PCR amplifications using primers targeting two regions of the ELVgv reverse transcriptase (RT), the three individual virus-host junctions and their corresponding pre-integration sites (proviruses* **A**, **B**, *and* **C***), and two control loci (FES, CHRNA1) in dermopteran genome (Janecka et al., 2008). CVO* - C. volans*; GVA3, GVA5* - G. variegatus *specimens; NTC - non template PCR control.*

We were successful in sequencing proviral parts from the 5’ end of the provirus from all three dermopteran individuals (GVA3, GVA5, CVO). Due to the limited amount and fragmentary nature of whole genome-amplified DNA, we were not able to obtain longer sequences than the ones depicted in Figure 18. We compared the obtained sequences to the sequence assembled from the GenBank data.

Due to the mutated nature of proviruses, we were able to amplify only four fragments of approximately 3 kb and shorter fragments of 0.7 – 1 kb in length in another four analyzed proviruses. 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. 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).

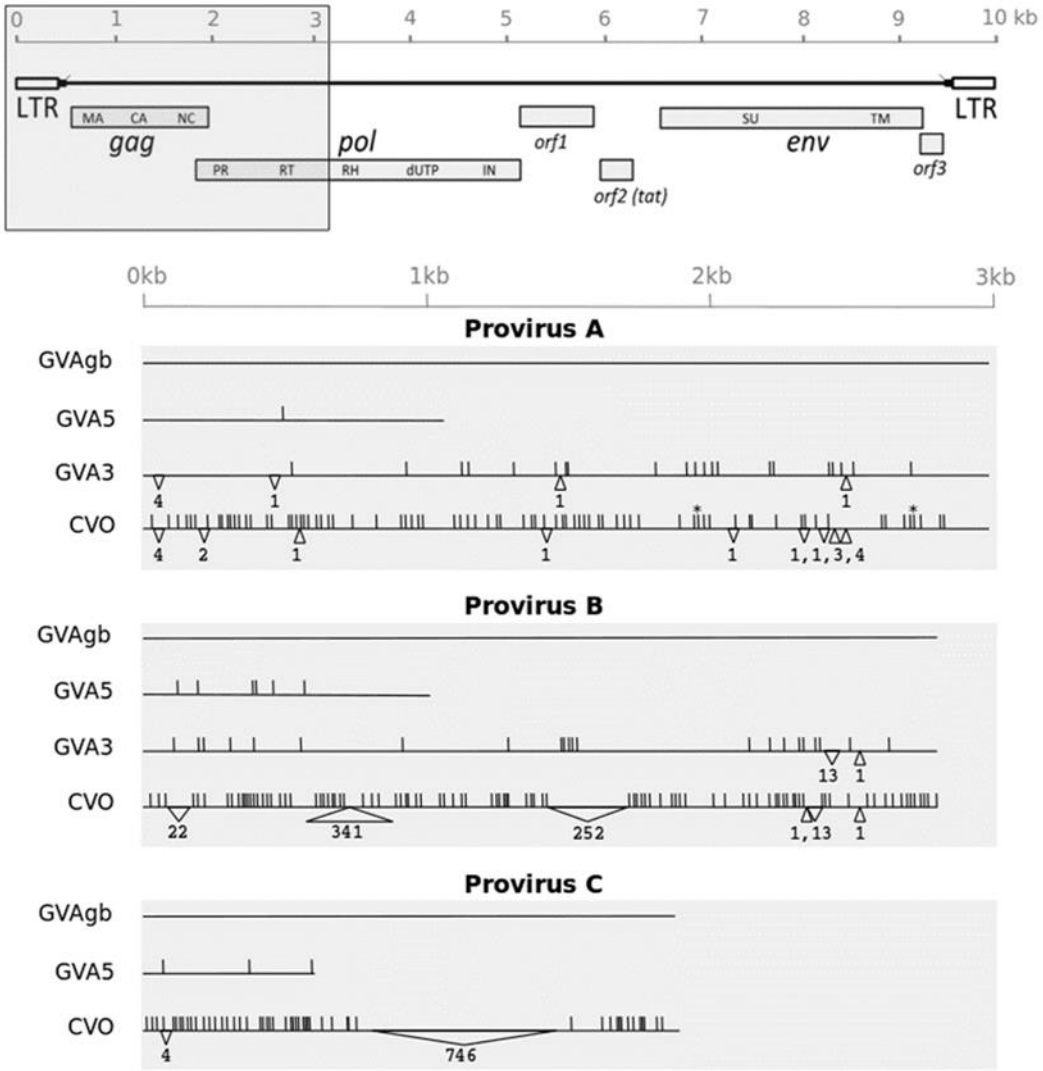

Figure 18: Sequence variability of three ELVgv proviruses integrated in the dermopteran genome. Upper: Schematic depiction of ELVgv genomic organization (Hron, et al. 2014). Region analyzed by sequencing is highlighted by grey box. Lower: Graphical representation of individual partial proviral sequences, with scale indicated above. 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.

The A, B and C proviruses, which represent independent integration events, are mutually different. Moreover, each provirus differs in many sites between individual animals. 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 further analyze the ELVgv evolutionary history in Dermoptera. This included more detailed analysis of ELVgv evolutionary history, confirming its position as the oldest lentivirus detected to date, and analysis of selection pressure acting at the antiretroviral restriction factor TRIM5α (Hron 2016).

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

Deltaretroviruses are an enigmatic genus of retroviruses. Up to this finding, no endogenous copy of a Deltaretrovirus was detected. In our screening efforts, we have detected remnants of endogenous Deltaretrovirus in the genome of Miniopteridae bats. We denoted this provirus MINERVa (Miniopteridae endogenous retrovirus).

We proved that this provirus is present in the *Miniopterus* genome only in a single copy, by thoroughful screen of the avaliable NGS and RNAseq data and by utilizing highly quantitative digital droplet PCR with primers designed to amplify LTR and *gag* sequences (see attached Manuscript).

We obtained samples from several bat species and confirmed their identity by sequencing their *cytB* or *RAG2* loci (see Table 5 for reference). The presence of MINERVa was confirmed by PCR detection using primers to amplify LTR (for detection of possible solo LTRs) and *gag* sequence, respectively. The MINERVa sequence was proven to be orthologous among all Miniopteridae specimens analyzed, but was not present in other analyzed bat species (see Figure 19).

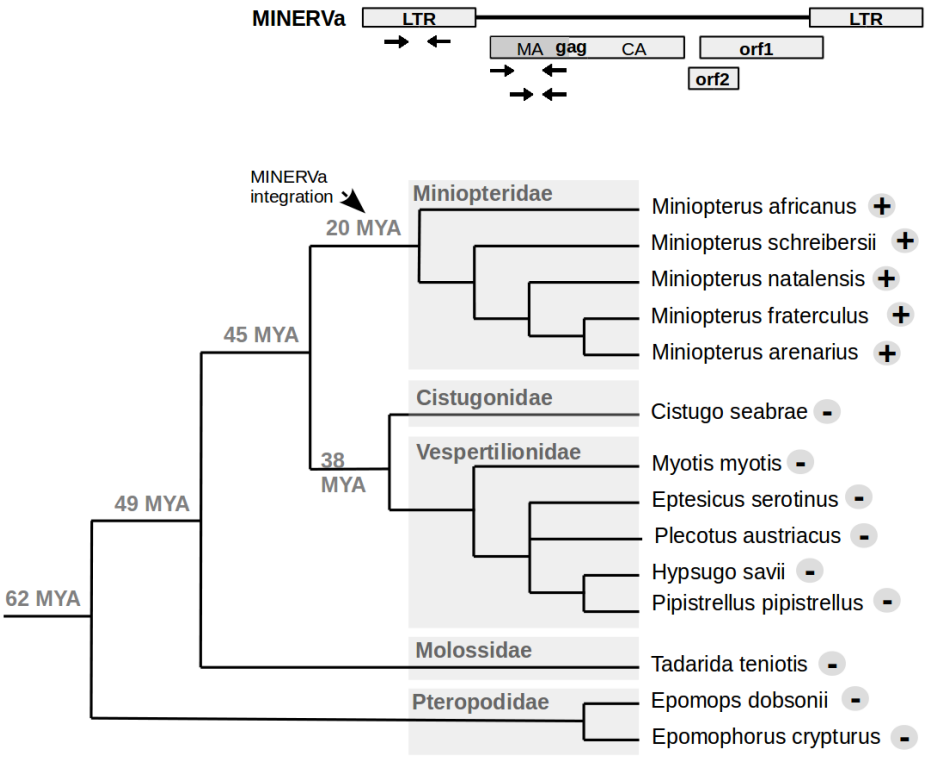

Figure 19: Absence/presence analysis of the MINERVa sequence in the bat samples. The time estimation at the nodes are the ones stated in TimeTree (http://timetree.org/).

MINERVa sequence was annotated by comparisons to two other well described deltaretroviruses (BLV and HTLV1). Splice site donor and splice site acceptor sequences were predicted using FruitFly (http://www.fruitfly.org/seq_tools/splice.html). Two additional ORFs were detected, which we presume to encode accessory genes (see Figure 20).

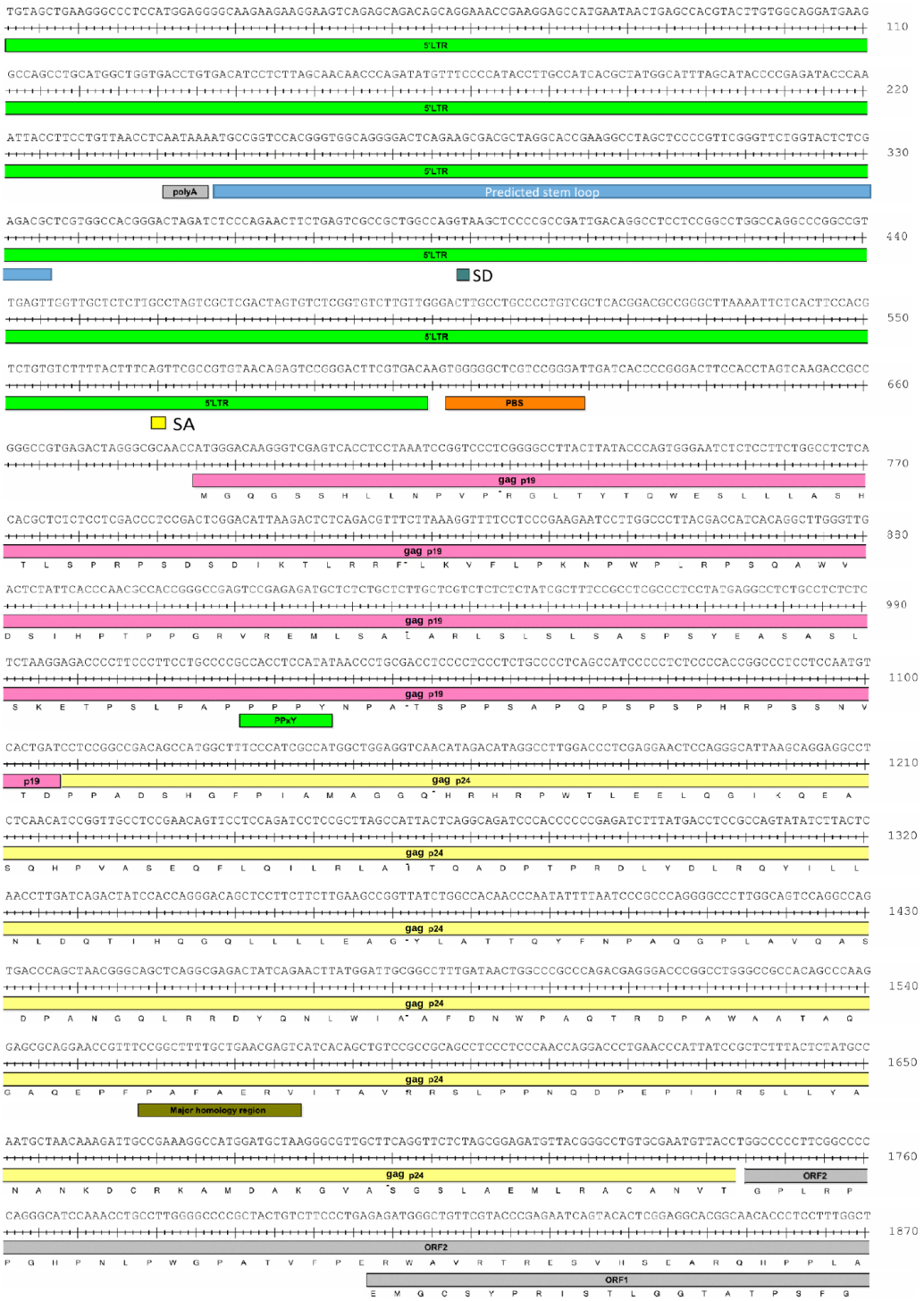

Figure 20: Annotation of the MINERVa sequence; PBS - primer binding site, p19 - matrix, p24 - capsid, polyA - polyadenylation signal site, ORF - open reading frame, SD - splice donor, SA - splice acceptor, PPxY - late doomain in gag.

Presence of the accessory gene coding for Rex in the genome was further indicated by prediction of stem loop structures characteristic for the Rex response element present in the LTR (see Figure 21). The predicted structure resembles the structures obtained by the same prediction approach in BLV and HTLV1.

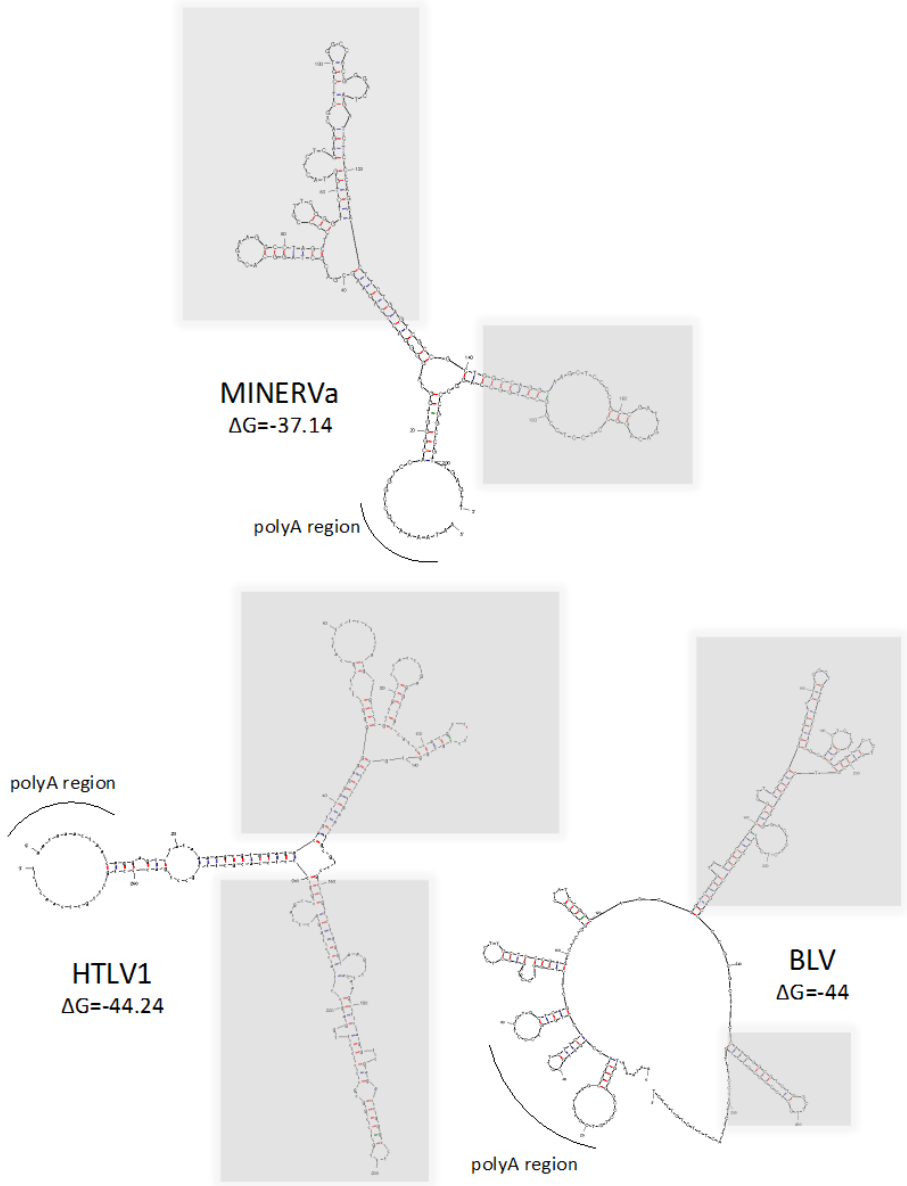

Figure 21: Prediction of a stem-loop structure in various deltaretroviruses using mfold software. The characteristic structures for the Rev Response element are highlighted in gray.

The presence of an intact ORF in the *gag* sequence was indicative of the fact that the provirus might produce a functional peptide sequence, leading to its purifying selection during Miniopteridae evolution. However, a pilot mass spectrometry analysis of the miniopterus muscle tissue did not yield a result confirming a presence of such a peptide. However, a muscle tissue is not optimal as a sample source in mass

spectrometry analysis, but was the only accessible at the time the analysis was performed.

This is the first report of a Deltaretrovirus possibly infecting a new mammalian order (Chiroptera). Interestingly, deltaretroviruses (namely BLV) are capable of replicating in a cell line derived from Tadarida species (personal communication with Kathryn Radke). Deltaretroviruses cause slow persistent infections in general, the viruses are predominantly spread via cell-to-cell transmission and if the immune cells are latently infected, clonal expansion occurs (Rafatpanah, Farid, Golanbar, & Jabbari Azad, 2006).

The deep evolutionary history of deltaretroviruses is not known. The sequences we detected and analyzed possess features of deltaretroviruses. Besides the mentioned sequential similarity to currently circulating deltaretroviruses, the sequence is cytosine-rich. It was previously shown, that HTLV nucleotide composition is biased due to being cytosine-rich and adenine-poor whereas the opposite applies to HIV (Kypr, Mrazek, & Reich, 1989).

Evidence of the presence of an endogenous Deltaretrovirus sequence is not only filling in the gap of knowledge of the last genus of retroviruses lacking in the currently available sequenced genomes, but might also help to understand the deep evolution of deltaretroviruses.

4.2 Molecular biology characterization of the currently endogenizing retrovirus (CrERV)

CrERV was recently studied due to its high insertional polymorphism among mule deer population in North America. However, almost all of the studies were only sequence-based or computational, and a virological characterization of the virus was lacking.

We were succesful in inducing the virus by cocultivation with susceptible human cells. The RT activity was measured in the coculture every week to monitor the start of the virus production. The normalized values of the RT activity and its growth during the analyzed time course is shown in the Figure 22.

This approach was based on the experiments performed upon the discovery of the inducible deer Gammaretrovirus, named DKV (Deer kidney virus) at that time (Aaronson et al., 1976). It also provided a source of the infectious virus at the first stages of the experimental work.

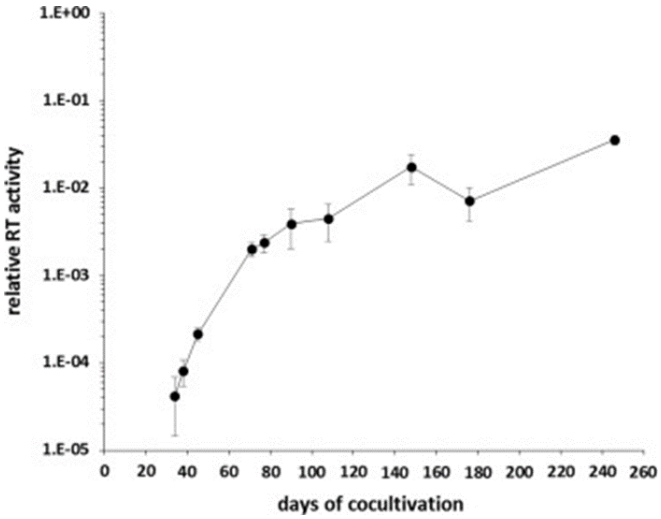

Figure 22: The relative RT activity of the medium on various days of cocultivation.

Because these experiments provided the first definitive virological identification of CrERV, and the titers obtained were extremely low, the particles of the induced CrERV were analyzed by various aproaches. First, we performed a gradient ultracentrifugation to test whether the virus sediments in the characteristic region of the Iodixanol gradient (See Figure 23). In sucrose density gradients, retrovirus particles sediment around 1.16

mg/ml (Contreras-Galindo et al., 2012). Iodixanol ultracentrifugation and visualisation of retroviral particles were performed on PERV in parallel. PERV and CrERV were used for comparison due to the fact, that PERV is widely studied in context of xenotransplantation and therefore a well characterized endogenous Gammaretrovirus (Kimsa et al., 2014).

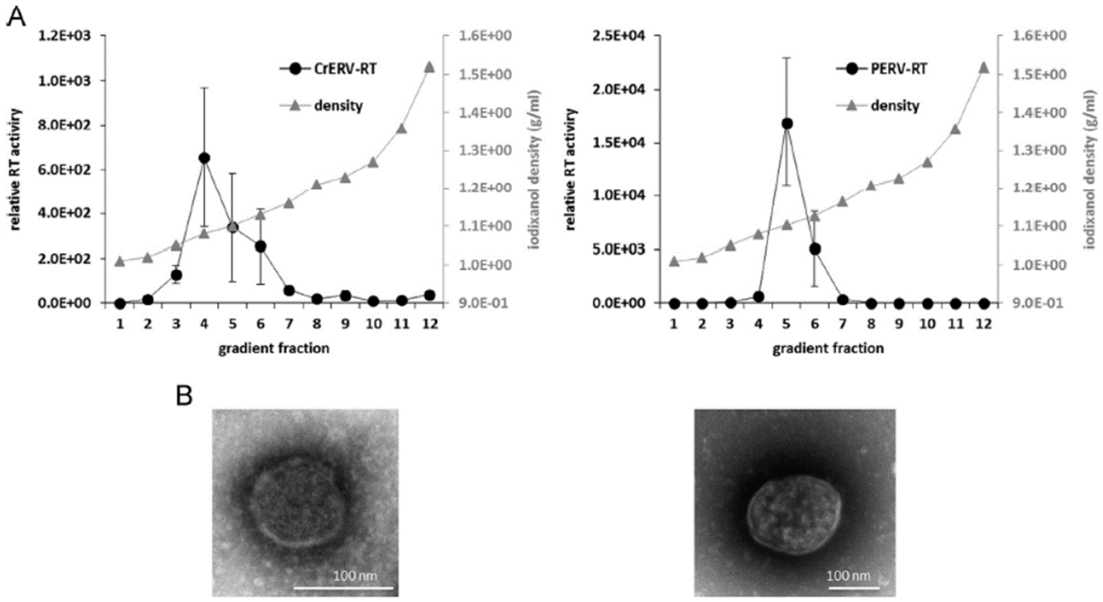

Figure 23: Presence of RT activity in Iodixanol gradient fractions for CrERV (left) and PERV(right) Comparisons of these fractions from two gradients confirmed, that the induced CrERV sediments in the fraction with density characteristic for retroviruses (A). Electron microscopy pictures of the retrovirus particles- CrERV (left) and PERV (right).

Second, we sequenced the full genome of the induced provirus and compared it to four closest full-lenght endogenous CrERV copies present and annotated in the mule deer genome (Kamath et al., 2014). As expected, the induced virus is the most similar to the endogenous copies estimated to be the youngest and differs the most from the proviruses with the oldest integration time estimates. This is in agreement with the assumption that among the hundreds of CrERV copies present in the mule deer genome, the evolutionary youngest copies have the greatest potential for induction and particle production. Based on the analysis of the obtained sequence of the induced virus and its comparison to other proviruses, the viral *env* presumably mutates at the fastest rate and the viral *pol* remains more conserved (Fig. 24). Based on the origin of the virus

(cocultivation by cells used by (Aaronson et al., 1976)), this comparison confirms the supposed identity of CrERV and DKV.

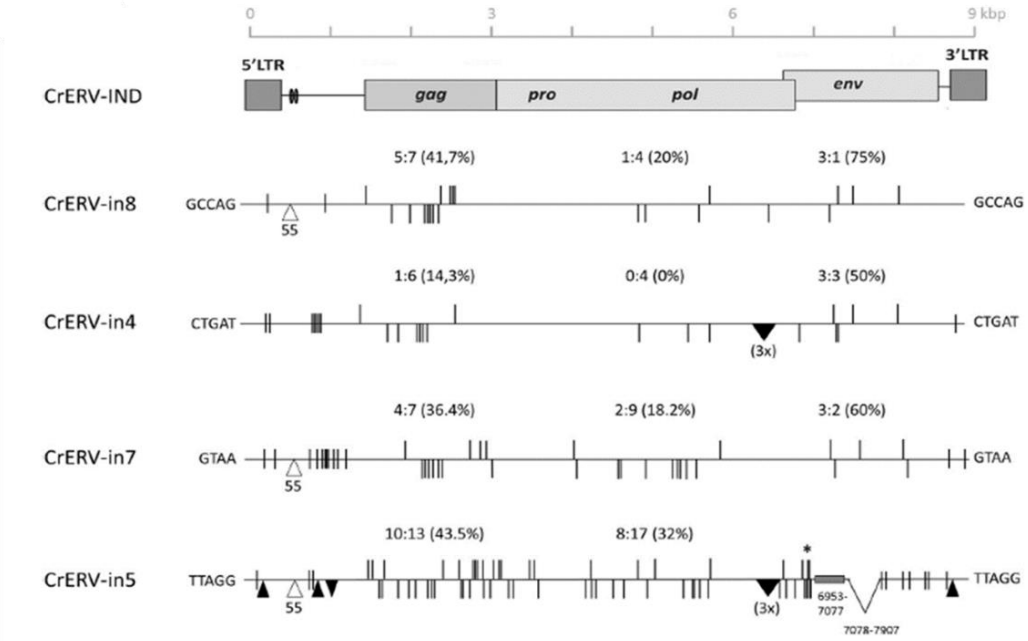

Figure 24: Comparison of full length seqences of the induced CrERV and its endogenous copies. The ticks representing the mutations in LTR and other noncoding viral sequences are aligned to center of the line representing each sequence. The ticks aligned above the line representing the sequence represents nonsynonymous mutations and the ticks aligned below the line represents the synonymous mutation. The ratio of mis-sense to sense mutation in individual genes is depicted above the individual lines representing the virus sequence. Insertions are depicted by triangles pointing towards the line, the deletions are depicted by triangles pointed outwards the line. Deletions or insertions in frame are depicted by black triangles, the frameshift indels are depicted by white triangles. The letters on both sides of the line represent the TSD sequence.

We further confirmed that the induced infectious virus (CrERV-IND) clusters with the evolutionary youngest copies in the mule deer genome by constructing a phylogenetic tree using 1 kb sequences from endogenous CrERV proviruses (Fig. 25). The extent of the 1 kb sequence was selected to minimize the impact of recombination events predicted in the various virus genomes.

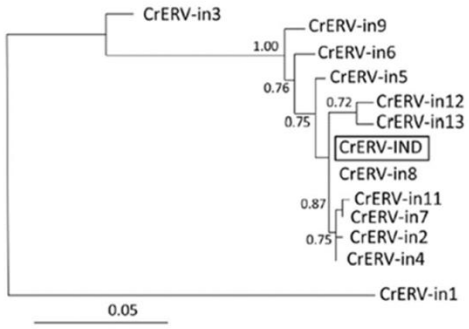

Figure 25: Phylogenetic analysis of the induced CrERV and other CrERV copies in the mule deer genome using 1 kb sequence located at the 3′ end of the provirus. The induced virus clusters with the evolutionary young copies.

The early and late infectious kinetics of the induced provirus was studied. The amount of newly made virus DNA (*env* gene product) of CrERV was estimated up to 48 h after infection. The experiment was performed in parallel with PERV and the infectious kinetics of the two viruses were compared (Figure 26). Early virus infection kinetics was shown to be regular, however only tiny fraction of cells was infected by CrERV (left); during long term infection of human cells, copy number around 1 viral DNA per cell was reached (right). The long-term spread of infection in the culture was also documented by repeated RT measurements.

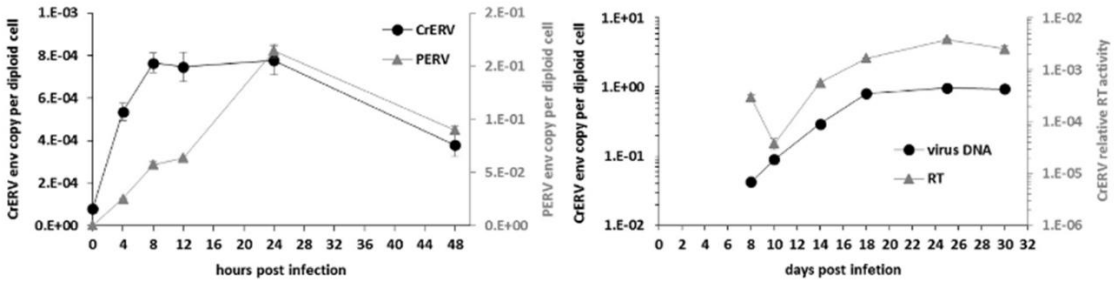

Figure 26: Early virus infection kinetics depicted (left), comparison of CrERV (circles) and PERV (triangles) env copy numbers per diploid cell up to 48 hours post infection; Long term infection of human cells, monitored up to 32 days post infection (right).

Despite being efficient in creating new endogenous copies, the virus shows xenotropic behavior (see Figure 31), in agreement with the previous observations of the Aaronson laboratory (Barbacid, Daniel, & Aaronson, 1980). For subsequent work, we prepared and tested an infectious molecular clone of the provirus. To be able to

genetically distinguish the endogenous CrERV proviruses and also the the newly made clone, we introduced four point mutation in the *pro* gene region to generate a mutant version of the molecular clone, CrERVmut (Fig. 27).

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 (Figure 27, lanes 7 and 8). The infectivity of the CrERVmut was confirmed on HEK293T cells (Figure 27, lanes 5 and 6). However, no viral DNA was detected when the CrERVmut virus was used to infect deer OHK cells (Figure 27, lanes 1 and 2). We therefore proposed the existence of a replication block in deer cells, occuring at a receptor level or a block occuring at an early stage of the retrovirus infection (virus uncoating or start of reverse transcription) blocking the efficient production of virus DNA. To analyze the capacity of CrERV to elicit a receptor interference, HEK293T cells were infected with wild-type exogenous CrERV and subsequently infected with CrERVmut. The cells chronically infected with wild type CrERV had close to one copy per cellular genome equivalent and presumably all expressed the virus envelope with the potential to block cellular receptors used for virus entry. The wild-type CrERV infected HEK293T cells did not exhibit production of CrERVmut virus DNA (Figure 27, lanes 3 and 4), in contrast with naive HEK293T cells. This is consistent with receptor interference being the cause of the resistance to CrERV on both deer cells and chronically infected human cells.

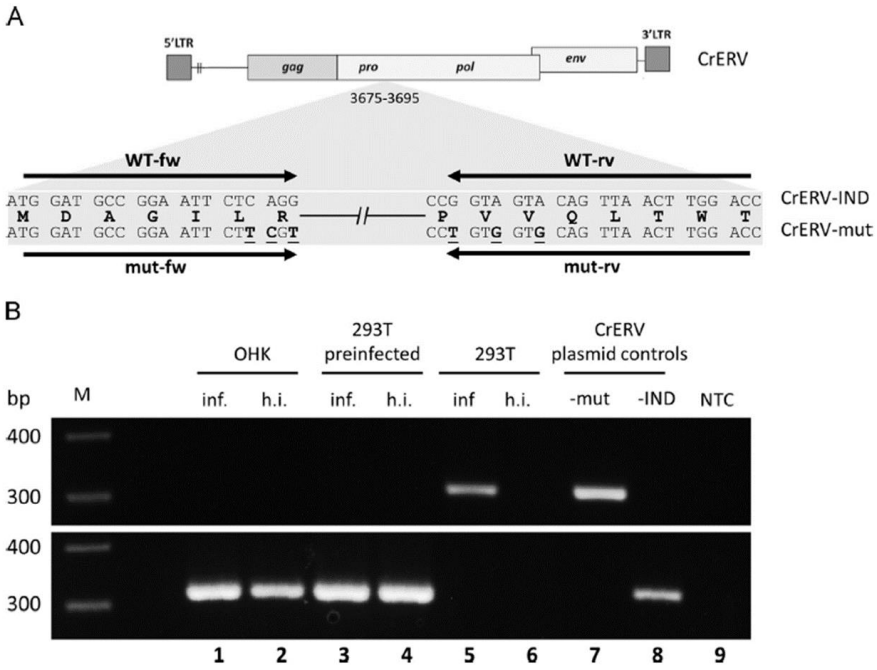

Figure 27: Construction of a CrERV molecular clone and its mutated form in order to distinguish endogenous and introduced copies of CrERV (A). CrERV-mut was used to infect deer OHK cells (lane 1), HEK 293 T cells (lane 5) and HEK 293 T cells chronically infected with CrERV-IND (lane 3). Heat-inactivated (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 speci fi city of PCR amplification. The upper panel shows PCR products generated with primers mut-fw and mut-rv, which detect speci fi cally the CrERVmut 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. (B).

The cellular receptor for CrERV is not known and the possibility that it might be mutated in deer remains unresolved. The mutation of CrERV receptor in mule deer might be an alternative explanation for the xenotropic behaviour of CrERV. Mutations of ERV receptors have been described in endogenous ALVs and MLVs (Barnard, Elleder, & Young, 2006; Kozak, 2014). The treatment of the cells with tunicamycin (inhibitor of N-linked glycosylation) has been shown to deglycosylate the cellular receptors or virus Env and thus enable the retrovirus entry by overcoming the receptor block (Koo, Parthasarathi, Ron, & Dougherty, 1994; Miller & Miller, 1992). However, tunicamycin treatment of both, deer and chronically infected cells did not rescue the cells susceptibility to virus infection (data not shown).

More variants of CrERV envelope genes may exist, utilizing different receptors and hence can overcome the interference blocks (Mary Poss, personal communication). This mechanism was described in FeLV and KoRV (Overbaugh et al., 2001; Xu, Gorman, Santiago, Kluska, & Eiden, 2015). More complex mechanisms were described for PERV, where disruption of a highly conserved PHQ motif in the N-terminus of Env enables transactivation of such viruses by unrelated gammaretroviral envelopes (Lavillette & 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 & Kabat, 2004). Interestingly, we observe a tendency toward disruption of the PHQ motif in the evolutionarily young CrERVs (data not shown).

Besides the phylogenetic analyzes of the CrERV induced by cocultivation, we performed further phylogenetic analyzes using a broad collection of gammaretroviral sequences. The accession numbers of the sequences are listed in the legend of the Figure 28. The ProtTest analysis of *pol* sequences of various retroviruses suggests that the closest endogenous retrovirus among the vertebrate endogenous retrovirus reservoir is an endogenous retrovirus in Greater horseshoe bat (*Rhinolophus ferrumequinum*). This result supports the notion that mammalian gammaretroviruses originated in bats (Cui et al., 2012). However, we did not include poorly characterized OERV (Ovine endogenous retrovirus) (Klymiuk, Muller, Brem, & Aigner, 2003) in this analysis, which could be evolutionary closer to CrERV than the aforementioned RfRV (data not shown).

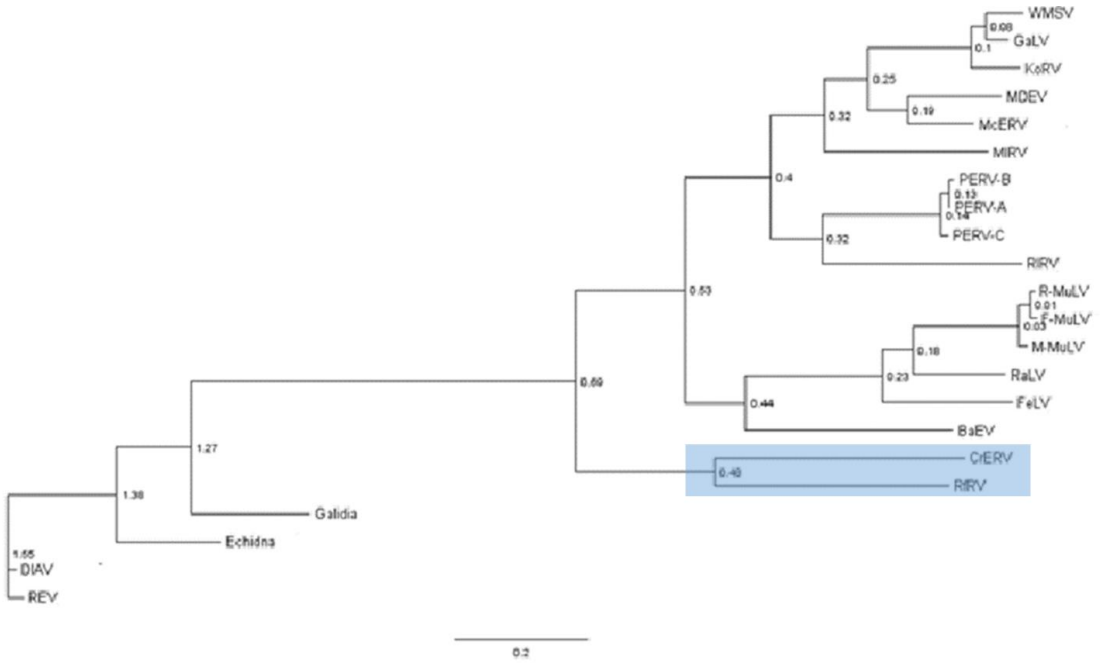

Figure 28: CrERV position in Gammaretrovirus phylogeny: Pol sequences were used for calculating the best fit tree to depict the protein evolution by algorithm used in Prot Test analysis; numbers displayed at nodes are the predicted age estimates calculated in ProtTest analysis; the accession numbers of the sequences used for the construction of the tree are: BaEV_BAA89659.1 (Baboon endogenous virus), FeLV_NP_955577.1 (Feline leukemia virus), F-MuLV_NP_040333.1 (Friend murine leukemia virus), GaLV_NP_056790 (Gibbon ape leukemia virus), KoRV_BAM67146.1 (Koala retrovirus), McERV_AGP25480.1 (Mus caroli endogenous retrovirus), MDEV_AAC31805.1 (Mus dunni endogenous virus), MIRV_AFM52260.1 (Megaderma lyra retrovirus), M-MuLV_NP_057933.1 (Moloney murine leukemia virus), PERV-A_AAM29192.1 (Porcine endogenous retrovirus A), PERV-B_AAM29194.1 (Porcine endogenous retrovirus B), PERV-C_CAC39617.1 (Porcine endogenous retrovirus C), RaLV_AAC78249.1 (Rat leukemia virus), REV_YP_223871.1 (Reticuloendotheliosis virus), RfRV_AFA52559.1 (Rhinolophus ferrumequinum retrovirus), RIRV_AFM52262.1 (Rousettus leschenaultii retrovirus), R-MuLV_NP_044738.1 (Rauscher murine leukemia virus), WMSV_YP_001165470.1 (Woolly monkey sarcoma virus), DIAV_AGV92859.1 (Duck infectious anemia virus), Echidna_ERV_AGV92856.1, Galidia_ERV_AGV92853.1, CrERV_AKA58521.1 (Cervid endogenous retrovirus).

4.2.1 Screening for CrERV polymorphism by pattern PCR

PCR between CrERV LTR and ubiquitous ungulate genomic SINE repeat was previously shown to yield pattern of bands, each corresponding to individual CrERV integration site (see Figure 29). This simple PCR-based assay can be used to screen for the presence of retrovirus integration site polymorphism, without the need to perform laborious NGS sequencing or repeated integration site cloning (Elleder et al., 2012).

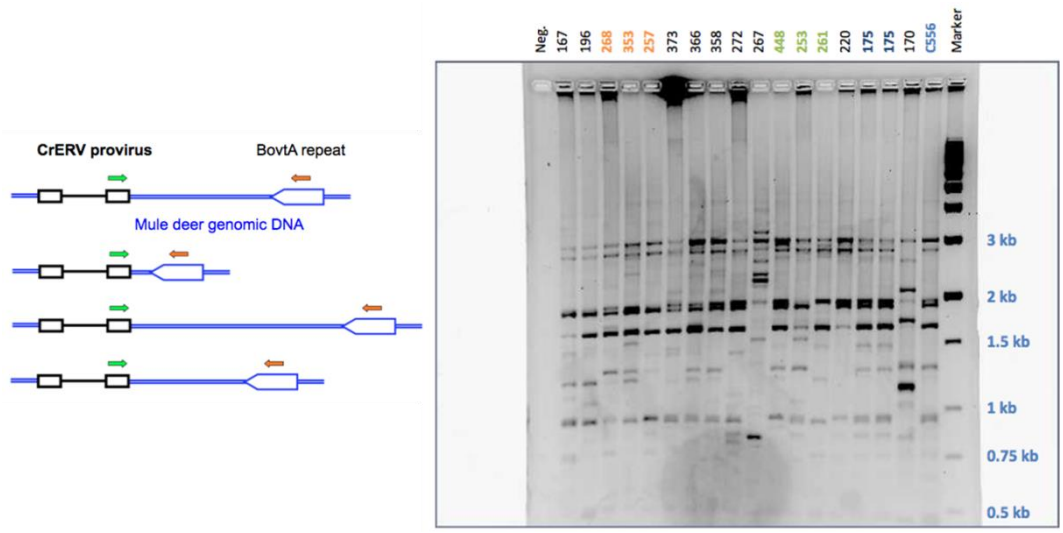

Figure 29: Strategy for determining the CrERV integration sites using primers targeting the abundant BovTA repeat and proviral LTR (left). Pattern PCR using these primers on various mule deer genomic samples (results of the further PCRs designed to amplify individual CrERV integrations are depicted in (Elleder et al., 2012).

By using the primers designed to amplify the BovTa repeat abundantly present in the Cervidae genomes, we tried to perform the pattern PCR to show the integration polymorphism of CrERV in other deer species (see Figure 30).

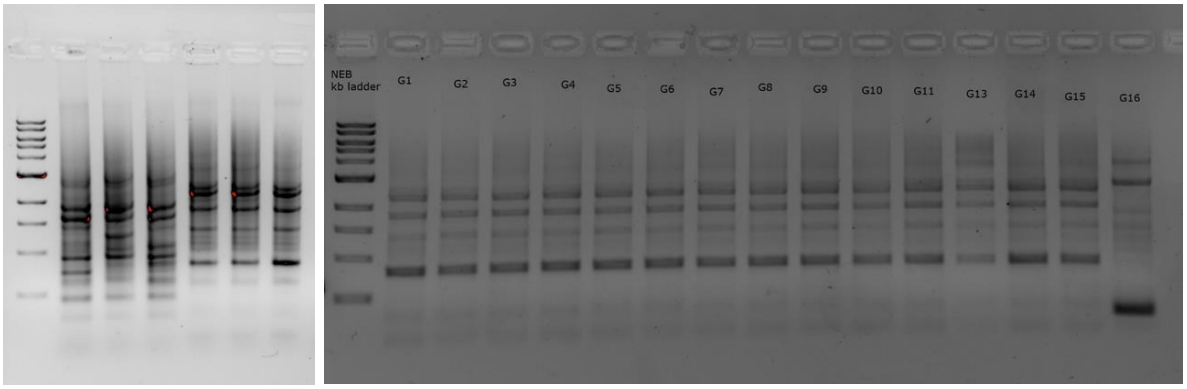

Figure 30: Pattern PCR from various Cervidae species. First lane is the molecular weight marker, next three lanes are various mule deer genomic samples and following three are various elk (Cervus canadensis) genomic samples (left ELFO). The right ELFO depicts the result of pattern PCR from various 15 red deer (Cervus elaphus) genomic samples (G1-15) and a genomic sample (G16) from European roe deer (Capreolus capreolus).

The pattern PCR assay performed did not distinguish between individual animals. However, the patterns of integration differs between species, indicating that BovTA repeat and CrERV were not present in the same loci in the common ancestor of the

analyzed Cervidae species. According to our current data, CrERV is present in all cervid species, however the extensive insetional polymorphism and continuing recent endogenization was only detected in the Odocoileus genus (mule deer and white-tailed deer (Elleder et al., 2012))

4.2.2 Further studies of CrERV tropism

As stated previously and in the presented work (Fabryova et al., 2015), CrERV exhibits xenotropic behavior. This finding is in contrast with the fact that CrERV is very efficient in creating new germ line copies, hence the high polymorphism among mule deer populations (Elleder et al., 2012). We tried to further characterize the replication blocks in CrERV infection, with emphasis on receptor-mediated stage of infection, using retroviral pseudotypes of CrERV and amphotropic MLV. We performed a marker rescue assay on permanently transfected human cells with CrERV using the VSV-G envelope (see the chapter 4.2.2.3 Marker rescue assays). The fact that titers of CrERV are extremely low caused large technical problems in reproducibility and conclusiveness of our results. Therefore, the results if this section are of preliminary nature and this topic is still actively persued in our laboratory.

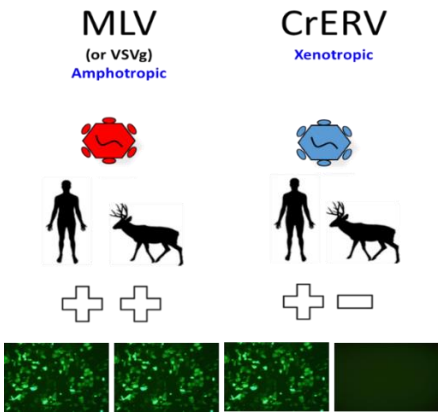

Figure 31: Infectivity of the amphotropic virus tested versus the infectivity of CrERV. Amphotropic virus infects both, the cells of the host as well as cells of other species, whereas xenotropic virus is not able to infect the cells obtained from the original host. The pictures of GFP positive cells are illustrative.

4.2.2.1 Retrovirus envelope pseudotypes

As mentioned above, in the experiments producing pseudotypes we struggled with a low titer of infectious particles produced. Some of the experimental results were

obtained by using flow cytometry, but due to low percentage of GFP-positive cells, the results could be biased by ambiguous setting of the threshold values.

Viral envelope-expressing constructs bearing a FLAG tag were also prepared to enable detection of particles by Western blot, and to check for levels of *env* expression and particle incorporation. Due to the very low *env* expression, the Western blots were not amenable to quantify the signals obtained. However, these vectors were proven to infect the cells more effectively than wild-type CrERV Env.

After many failed attempts to detect the percentage of GFP-positive cells using flow cytometry, we tried to observe the foci of green cells under fluorescent microscope 3 dpi. The first three infections were performed in weekly period on a 100 mm tissue culture dish yielding the following results:

Experiment 1: The cells were infected with AmphoenvFLAG construct, CrERVenvFLAG construct produced by transfection of HEK293T cells by according plasmids. For the reference about the experiment set up see Table 9. Transfection of sole *gag-pol* construct was was used as negative control in infectivity. The infection with construct with AmphoenvFLAG yielded 24 GFP-positive foci and construct with CrERVenvFLAG yielded 20 foci after infection of HEK293T cells.

Experiment 2: This experiment was a replication of the Experiment 1. Transfection of sole *gag-pol* construct was inefficient; the infection with construct with AmphoenvFLAG yielded 48 foci and construct with CrERVenvFLAG yielded 15 foci after infection of HEK293T cells.

Experiment 3: The aim of this experiment was to repeat Experiments 1 and 2 and try to increase the number of yielded GFP-positive cells by using RetroNectin. The infection of sole *gag-pol* construct yielded no GFP positive cells; the infection with construct with AmphoenvFLAG yielded 75 foci and construct with CrERVenvFLAG yielded 5 foci after infection of HEK293T cells. We tried to enhance the effectivity of infection using RetroNectin (see the chapter 3.2.8 Infection for reference about the mechanism of action) using the spin procedure according to manufacturer's protocol.

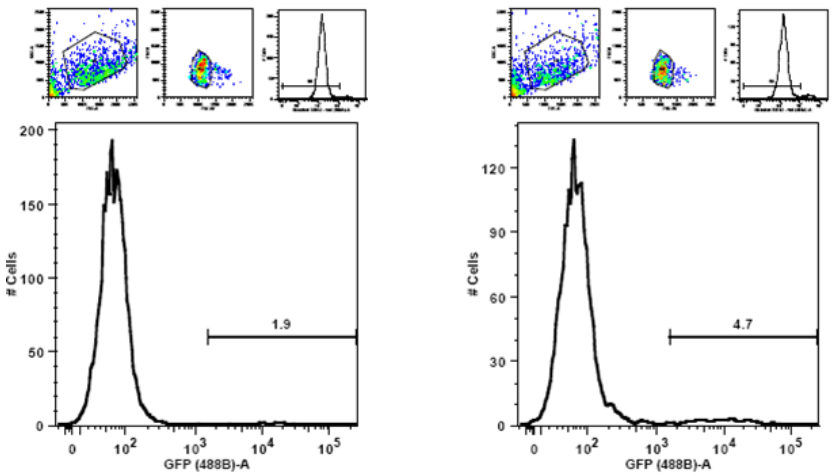

Figure 32: RetroNectin treatment of the tissue culture plates increases the number of infected cells by an amphotropic virus. The results of FACS analysis of infection without the use of RetroNectin (left) and with RetroNectin (right) indicate the more than two fold increase.

We examined the potential of elevating the number of infected cells by using RetroNectin using medium from the cells stably producing a virus with amphotropic properties. The usage of RetroNectin elevated the number of infected cells more than two-fold (see Figure 32 for reference).

Experiment 4: The RetroNectin-coated 6-well plates with 4 ml of produced virus spun on them were used. Cells HEK293T were seeded in 10% confluence and green foci were counted 3 dpi. The usage of RetroNectin did not significantly increase the number of GFP-positive cells after infection with a construct with an amphotropic Env. However, it was shown to increase the number of GFP-positive cells afer infection with a construct CrERVenvFLAG (see Table 10).

Table 10: Increasing the virus infectivity on HEK 293T cells by coating the dishes with RetroNectin. The numbers represent the number of GFP-positive cell foci on the analyzed dishes.

Construct	**4ml of virus per well Without RetroNectin**	**4ml of virus per well With RetroNectin**	**2ml of virus per well With RetroNectin**
Gag-pol only	0	0	0
AmphoenvFLAG	8	11	10
CrERVenvFLAG	2	19	4

Experiment 5: Juxtaposition of infectivity of the pseudotyped virus on HEK293T cells and and HEK293T cells bearing a copy of CrERV (denoted Cr5). The number of cell foci is counted on day 3 and on the day 4 (depicted as number of foci at day3/ number of foci at day4). What came as a surprise in this experiment was the fact, that Cr5 cells did not exhibit a decreased sensitivity to a pseudotype with a CrERVenv compared to HEK293T cells. This result was never replicated.

Table 11: Juxtaposition of infectivity of the pseudotyped virus on HEK293T cells and and HEK293T cells bearing a copy of CrERV (denoted Cr5). The number of cell foci is counted on day 3 and on the day 4 (depicted as number of foci at day3/ number of foci at day4).

Construct	**Cr5**	**HEK293T**
Gagpol only	ND	0/0
CrERVenvFLAG	7/9	4/9
AmphoenvFLAG	9/10	59/uncountable

Experiment 6: Juxtaposition of infectivity of the pseudotyped virus on HEK293T cells and and HEK293T cells bearing a copy of CrERV (denoted Cr5). RetroNectin was used this time, but cells were seeded at a low confluence (15,000 cells/well in 6 well plate). The number of foci was counted 7 dpi. Due to the result from the previous experiment, we infected Cr5 cells also with the medium from C10 cells (HEK293T cells producing CrERV and transfected with a plasmid with GFP). The Cr5 cells were shown not to be susceptible to either, infection by CrERVenvFLAG pseudotype nor C10 supernatant (see Table 12).

Table 12: Comparison of HEK293T cells susceptibility and HEK293T cell line chronically infected with CrERV construct

Infected by	**Cr5 cell line**	**HEK293T**
C10 supernatant	0	17
CrERVenvFLAG	0	2
AmphoenvFLAG	12	2

Experiment 7: The previous experiments were repeated. This time the flow cytometry analysis to estimate the percentage of GFP positive cells was used. Due to low percentage of GFP-positive cells, the results could be biased by ambiguous setting of the threshold values. All of the cells infected with the medium from cells transfected only with *gag-pol* and pLG plasmid (to control possible contamination with GFP-positive cells

from the transfected cells) examined by eye were considered negative. However, these measurements indicate that HEK293T cells are infected by constructs with CrERVenvFLAG approximately ten times more effectively than Cr5 cells (see Table 13).

Table 13: Percentage of GFP positive cells after infection

Construct	**HEK293T**		**Cr5**	
	RetroNectin	w/o RetroNectin	RetroNectin	w/o RetroNectin
Gag-pol MLV only	0.002	0.0026	0.0064	0.0016
CrERVenvFLAG	0.05	0.035	0.0016	0.0037
AmphoenvFLAG	0.033	0.017	0.014	0.014

Experiment 8: After obtaining cells from two different cervid species, we infected them with our pseudotype constructs. We used 10 cm plates and infected the cells with 12 ml of the prepared construct. The foci were counted 7 dpi. Cells were fixed (see chapter 3.2.12 Cell fixation for DAPI staining) and the pictures were taken using fluorescent microscope (see Table 14).

Table 14: Infecting deer species and human cells with various viral constructs

Cells infected	**Gag-pol**	**CrERVenvFLAG**	**Ampho**
RED DEER	0	0	1
FALLOW DEER	0	0	0
OHK	0	0	3
HEK 293T	0	0	10

This result indicate that the replication block is not present only in mule deer cells (OHK) but presumably also in cells obtained from other Cervidae species (see Table 14). Cells from Cervidae species were proven to be less sensitive to infection by both CrERVenvFLAG and AmphoenvFLAG constructs than HEK293T cells (see Figure 33).

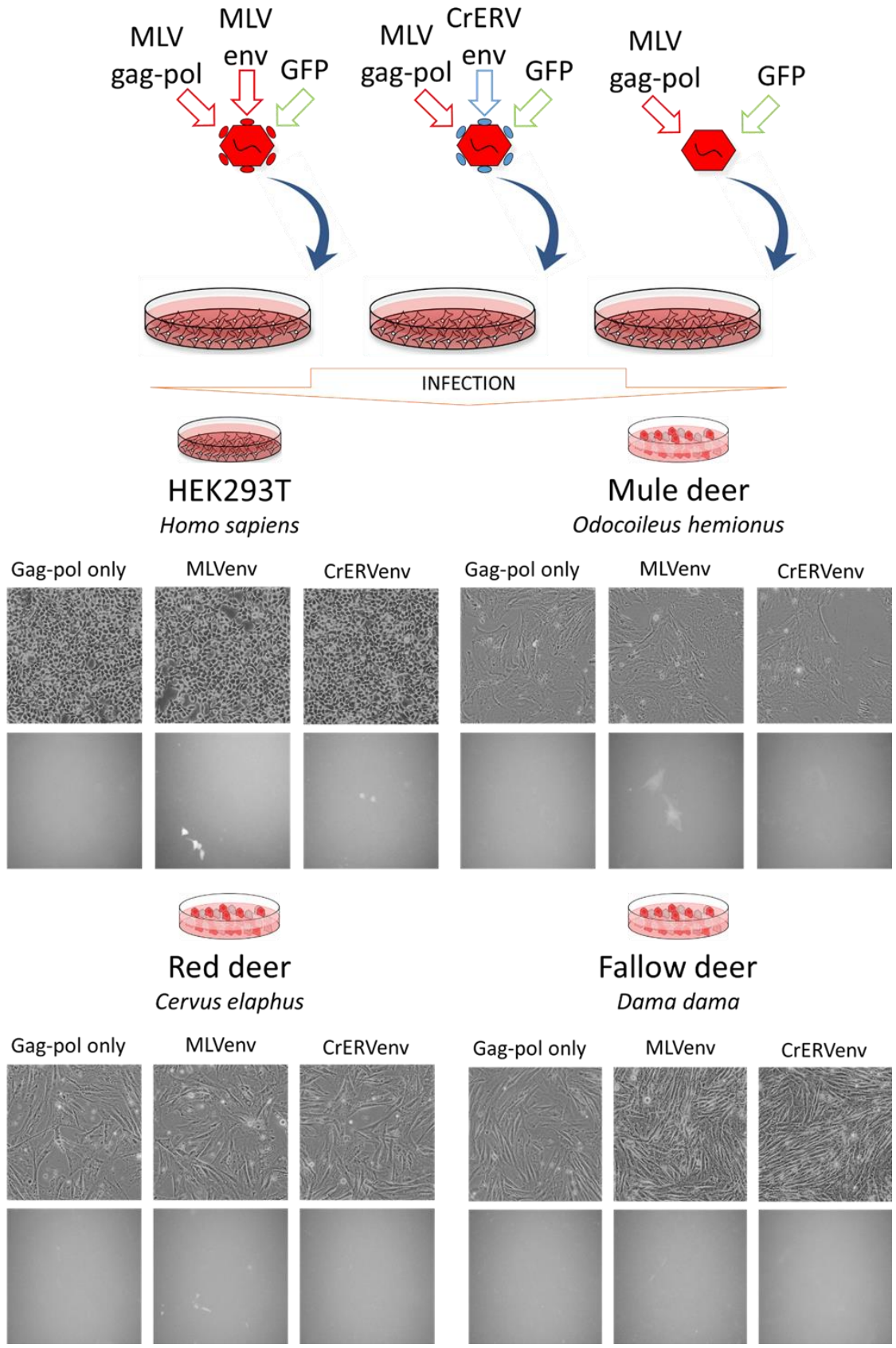

Figure 33: Outcome of the pseudotyping experiment 8. The cells photographed in bright field are displayed in the upper rows, the capture of the GFP positive cells is displayed in the lower rows.

Despite the fact that we repeatedly confirmed that the constructs with MLV envelope are able to infect mule deer, red deer, and fallow deer cells whereas the constructs with CrERV envelope are not, we always detected only a small number of GFP-positive cells in the experiment set up with the susceptible cells. HEK293T cells were generally shown to be more susceptible to infection. After repeated attempts and only minor success in elevating the percentage of GFP-positive cells to enable us to provide more rigid data, we started using the marker rescue assays as an alternative approach.

4.2.2.2 Increasing the effectivity of the CrERV by removing the cytoplasmic domain of the env

In order to increase the effectivity of the infection, we constructed a plasmid with *env* lacking the cytoplasmic domain (aforementioned envnT in methods). This approach was proven to enhance the effectivity of the infection in SIV (Kuwata, Kaori, Enomoto, Yoshimura, & Matsushita, 2013), HIV (Ye et al., 2004), and enhance specifity in immunoblotting assay in HTLV1 and HTLV2 detection (Varma et al., 1995). However, we were not successful in elevating the efficiency of the infection sufficiently to enhance the quality of the outcomes of the pseudotyping experiments.

4.2.2.3 Marker rescue assays

For the first experiment with marker rescue assay, 2 ml of fresh C10 cells medium was used for infection or 1 ml of medium of CrERV positive cells transfected with VSV envelope. Three days after infection, cells were fixed and stained with DAPI to yield highly fluorescent nuclei and no detectable cytoplasmic fluorescence. After two weeks, PERT assay was performed to estimate the RT activity in the medium.

After infecting Cervidae cells with medium from CrERV positive cells, no GFP signal was detected under fluorescent microscope (data not shown) contrary to infecting HEK293T cells. After infecting both Cervidae and human cells with medium from CrERV positive cells transfected with VSV-G envelope GFP signal was observed (see Figure 35). The medium from cells cotransfected with GFP plasmid, MLV *gag-pol* and VSV-G *env* was used as a control. However, the cotransfection did not produce replication competent virus, so no reverse transcriptase activity was detected in the medium two weeks post infection (see Figure 34).

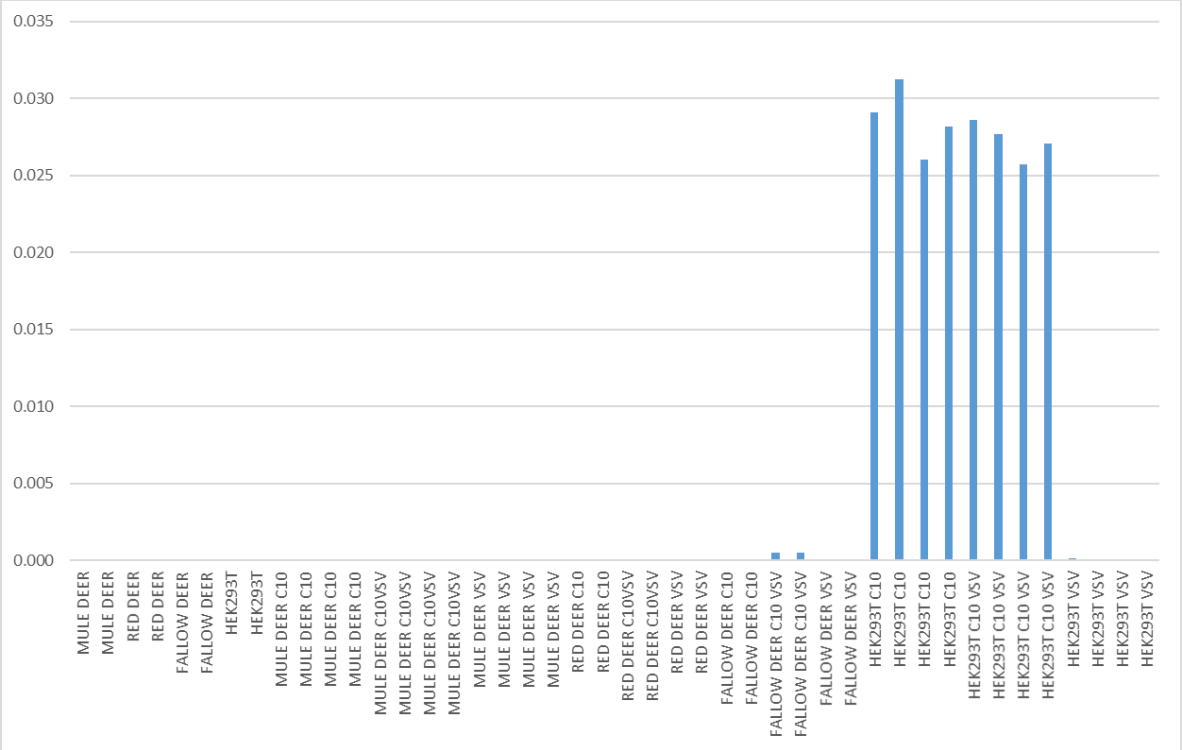

Figure 34: PERT assay after two weeks post infection with CrERV rescued VSV env. The experiments were performed in a technical duplicates and tetraplicates. The numbers on the y axis represent the relative RT activity. The individual columns (from left, duplicates) represent uninfected mule deer cells, uninfected red deer cells, uninfected fallow deer cells, uninfected HEK293T cells. The mule deer cells infected with medium from C10 cells (tetraplicate), the mule deer cells infected with medium from C10 cells transfected with VSV envelope (tetraplicate), with medium from C10 cells (tetraplicate), with construct with MLV gag and pol and VSV env unable to replicate itself (tetraplicate). The red deer cells and fallow deer cells were infected in duplicates in the same fashion. The HEK293T cells were infected in the same manner as mule deer cells in tetraplicates.

The marker rescue is an alternative approach to pseudotype experiments. The infection by a medium from C10 cells transfected with a VSV-G plasmid yields GFP-positive cells. The medium from HEK293T cells co-transfected with an MLV *gag-pol* plasmid, VSV *env* and a plasmid with a GFP signal yields a number of GFP positive cells comparable to the number obtained after infection with the rescued virus.

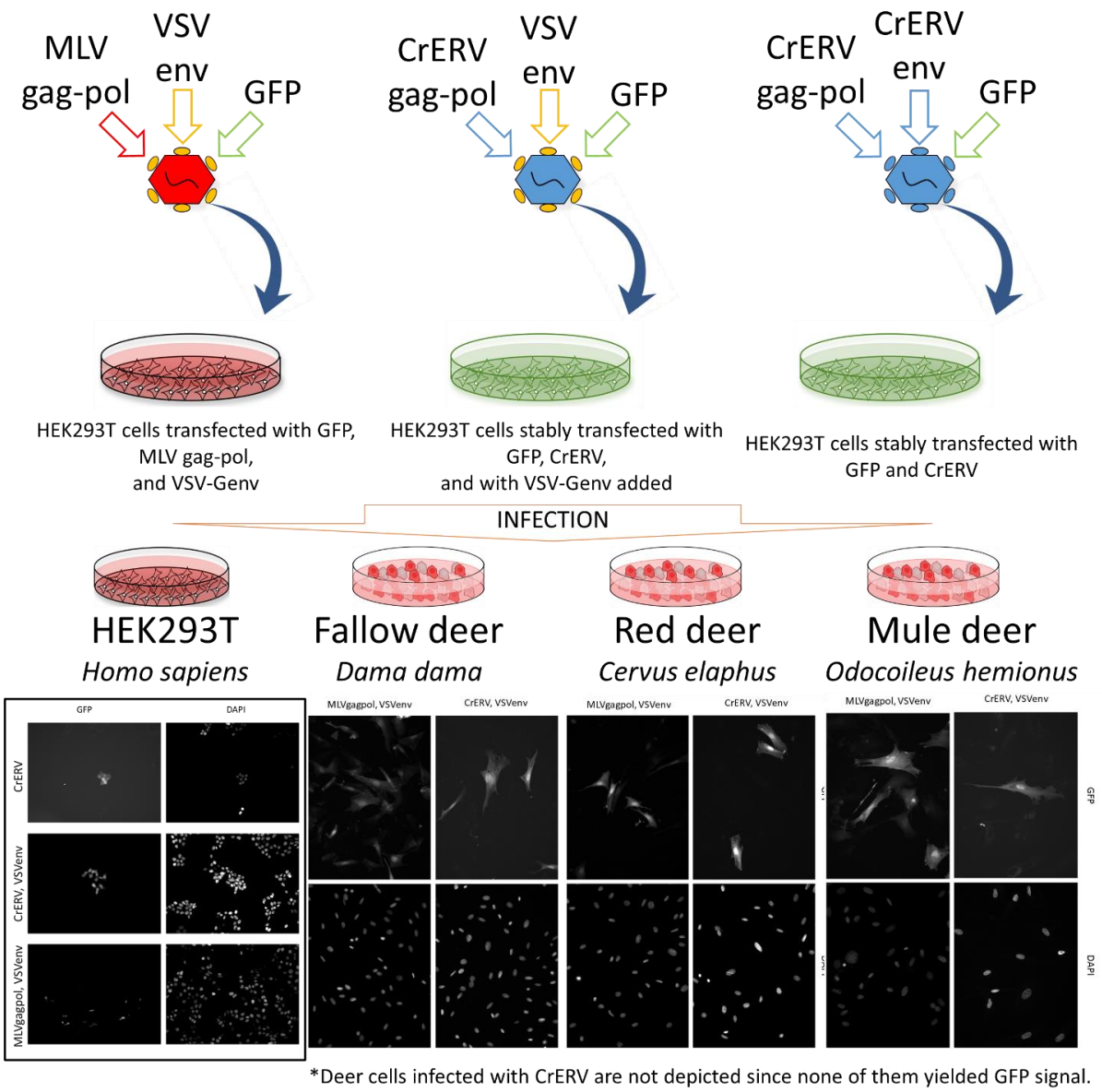

Figure 35: The marker rescue experiment - 3 days post infection. Cells used for infection are listed next to the row, the virus construct strategy is denoted above the pictures. The outcomes of infection by C10 medium are not displayed for fallow deer, red deer, and mule deer cells for they were all negative.

4.2.2.4 Late block of CrERV infection

Based on the previous results, we concluded that a late block of replication might be present in CrERV infection of deer cells. We therefore transfected the cells using Cell Line Nucleofector Kit (LONZA, Switzerland), and thus made sure that virus is delivered directly to the nucleus. RT activity in the medium using PERT assay was measured after transfection, three days post transfection, six days post transfection, and seventeen days post transfection. However, we detected RT activity only in the medium of HEK293T cells after transfection and not in the mule deer, fallow deer and red deer cells. Three days post transfection we did not detect the RT activity even in the medium of HEK293T cells. Six and seventeen days post transfection all of the transfected cells (or rather the medium in which they were cultivated) remained negative.

These preliminary results need further validation and additional experiments are needed.

4.2.3 Host restriction factor assembly and analysis

4.2.3.1 APOBEC

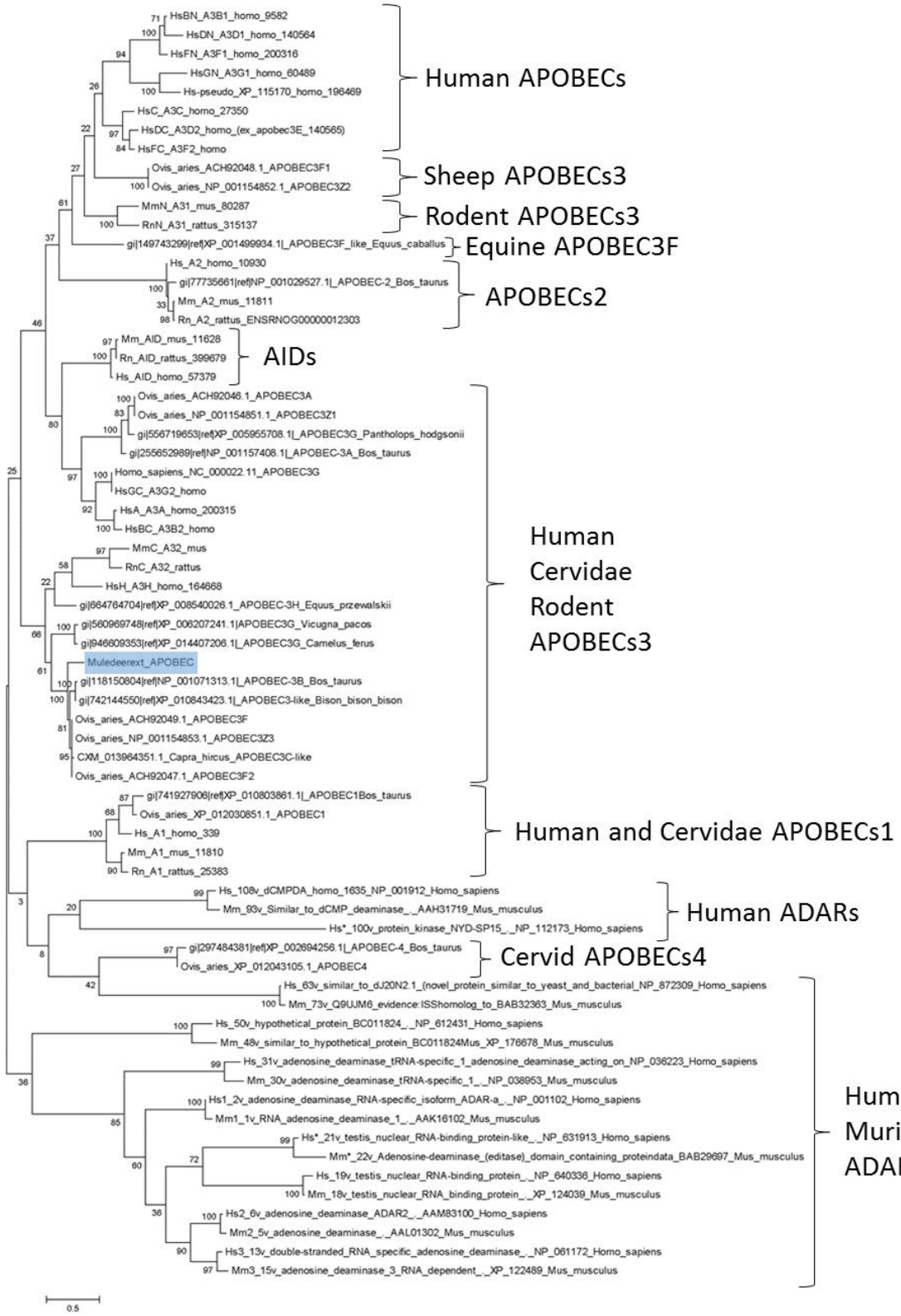

Figure 36: The Neighbor joining tree using Muscle alignment of APOBECs available in NCBI nucleotide database (accession numbers displayed in the tree) and APOBECs, AIDs, and ADARs

listed in (Conticello, Thomas, Petersen-Mahrt, & Neuberger, 2005). Mule deer APOBEC assembled from NGS data is highlighted in blue.

We attempted to assemble *APOBEC3* gene using the mule deer NGS data. However, the data available yielded low coverage in some areas of the gene. Because APOBEC isoforms are poorly described in Cervidae species (see Figure 36) we were unable to neither determine nor assemble precisely one particular isoform of the APOBEC. Hence, we were unable to perform analysis of positive selection for it would be greatly biased.

4.2.3.2 SAMHD1

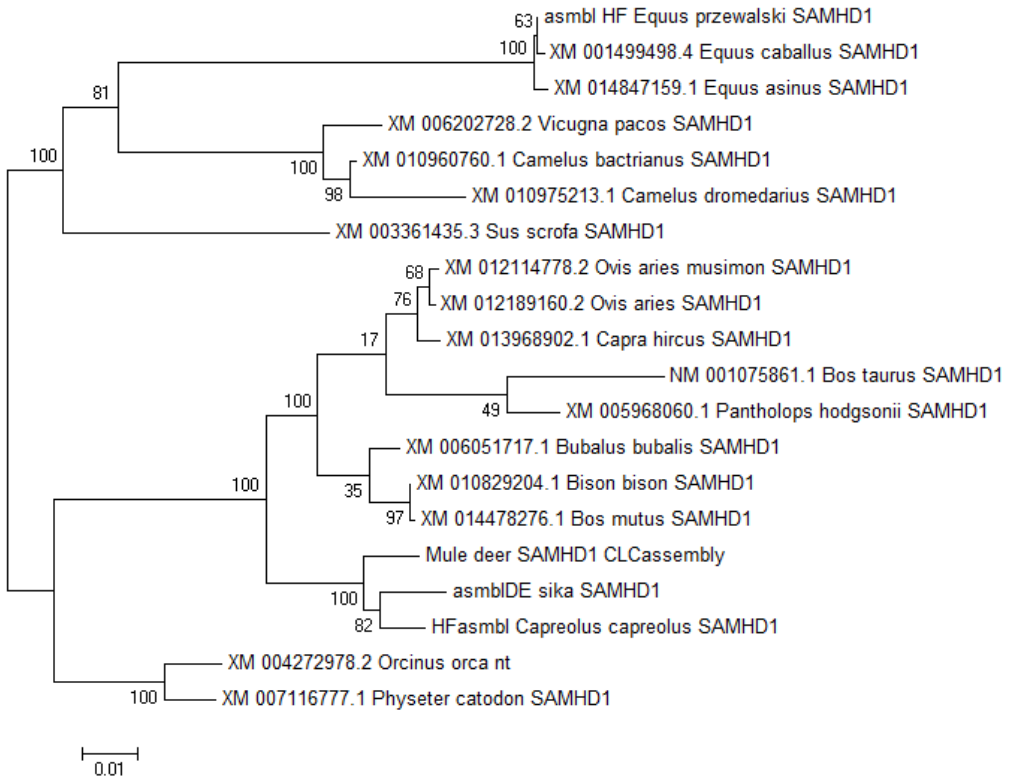

Figure 37 The Neighbor joining tree using Muscle alignment of SAMHD1s available in NCBI nucleotide database (accession numbers displayed in the tree) and SAMHD1 sequences assembled in CLC Workbench (denoted as asmbl).

We assembled SAMHD1 in the mule deer using SRA data from lymph nodes. SAMHD1 was assembled in additional species, being sika deer (*Cervus nippon*), Przewalski's horse (*Equus przewalski*), and European roe deer (*Capreolus capreolus*). We constructed a phylogenetic tree using SAMHD1 sequences avaliable at NCBI. The

obtained sequences cluster with related species with high bootstrap supports (shown at the nodes in Figure 37).

4.2.3.2.1 Branch-site REL (random effects likelihood) analysis

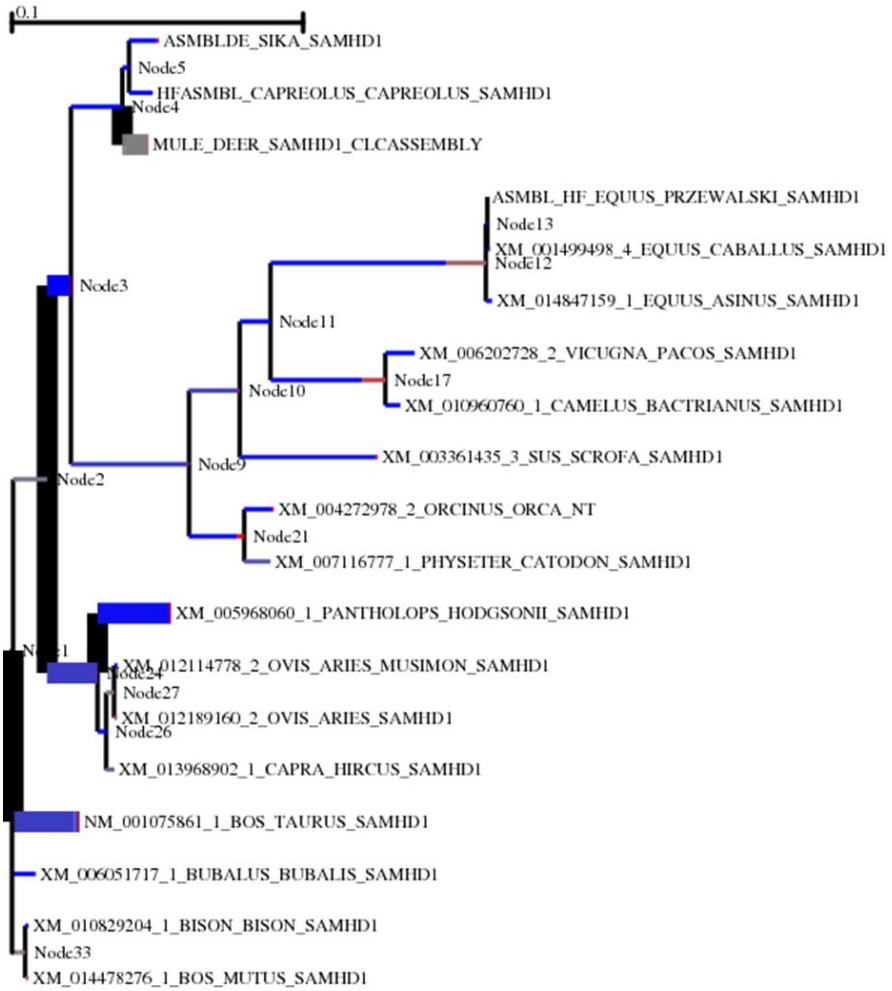

Figure 38: Branch site REL analysis of the SAMHD1 gene. The accession numbers of the used sequences are depicted in the branches. Sika, Equus przewalski, mule deer, and Capreolus are de novo assembled sequences of the sika deer (Cervus Nippon), Przewalski's horse (Equus przewalski), mule deer (Odocoileus hemionus), and European roe deer (Capreolus capreolus).

The obtained sequences were analyzed in respect to possible presence of their positive selection. The results are depicted in the Figure 38. The thickness of the horizontal lines represent p-values calculated for each branches, the thickness of vertical lines depicts the p-value calculated for the nodes. The thick line depicts the p-value of less than 0,05. The length of the branch displays the number of nucleotide changes compared to the neighbor sequences. The hue of each color indicates strength of selection, with primary red corresponding to dN/dS >5, primary blue to dN/dS = 0 and

grey to dN/dS = 1. The width of each color component represents the proportion of sites in the corresponding class (Kosakovsky Pond et al., 2011).

Hence we can state the most SAMHD1 nucleotide sites of mule deer is under neutral selection based on the dN/dS calculation with significant p-value calculated for the analyzed branch. However, some nucleotides under positive selection (dN/dS >5) are present, but they are low in number and their presence is not eligible from the presented tree (see Figure 38).

4.3 Analysis of the infection block in CHOK1 cells

Mounting evidence supports the concept of products (or secreted products) of some endogenous retroviruses blocking infection by particular exogenous retroviruses (Malfavon-Borja & Feschotte, 2015). CHOK1 cells are not susceptible to infection by exogenous retrovirus with an amphotropic envelope. Introduction of human amphotropic or mouse ecotropic retrovirus receptor or tunicamycin treatment of the cells was shown to rescue the aforementioned susceptibility. However, at the time of these findings, current methods were not available and the exact mechanism of the resistance was not explained. The sequence of the CHOK1 genome (Lewis et al., 2013) as well as availability of the proteome (Baycin-Hizal et al., 2012) became accessible only recently and the cells are still being characterized from various aspects (Hefzi et al., 2016).

We were successful in rescuing the CHOK1 susceptibility to amphotropic MLV vector by tunicamycin treatment (see Table 15), which is probably modifying the cell receptor properties. Importantly, previous reports indicate inhibitory effect of the medium conditioned by CHOK1 cells and detect the presence of secreted inhibitors of retrovirus infection of protein nature. This protein was proven not be an interferon by replicating the infectivity experiments on CHOK1 deficient in the interferon production (Miller & Miller, 1992). The inhibitory agent is presumed to be a protein secreted by hamster cell lines; it is also present in the Syrian hamster (*Mesocricetus auratus*) and Chinese hamster (*Cricetulus griseus*) serum. Mutation in the glycosylation pathway was proven to increase the susceptibility of CHOK1 cells to the amphotropic MLV (Miller & Miller, 1993).

Table 15: Reproduction of the infectivity experiment using CHOK1 cells performed by (Miller & Miller, 1992, 1993). The conditioned medium column depicts the addition (+) or absence (-) of the medium in which CHO cells were grown for 24 hours. This medium contains the studied secreted inhibitors (Miller & Miller, 1992). For details about the infections see the chapter 3.3 Methods related to the CHOK1 cells project.

Tunicamycin treatment	**Conditioned medium**	**RetroNectin**	**GFP positive cells (%) 3 d.p.i.**
+	-	-	1.76
+	-	+	0.016
+	+	-	0.012
+	+	+	0
+	+ (without serum)	-	0.42
-	-	-	0.057
-	-	+	0.01

As previously displayed (see Figure 32) the usage of RetroNectin elevated the number of infected cells more than two-fold. However, RetroNectin did not increase the number of the infected cells in this experimental setup (Table 15), possibly due to the fact that it might interfere with the deglycosylation of the receptors mediated by Tunicamycin. We also tested the inhibitory activity of the conditioned medium without serum; the use of serum-free medium is necessary for our planned mass spectrometry analysis. The presence of the serum was shown not to interfere with the block to virus infection (see Table 15, lane 5). However, depletion of the serum decreased the inhibitory activity of the conditioned medium to some extent. This might be due to the fact that the cultured cells were deficient for some nutrients for 24 hours, which might lead to the decreased secretion of the inhibitory proteins into the medium.

After replicating the previously reported experiments, we concentrated and fractionated the conditioned medium by gel filtration. Thirty-two fractions of the medium were harvested. The secreted inhibitors were reported to have molecular weight between 10-50 kDa (Miller & Miller, 1992). These expected molecular weights were supposed to be present in fractions 9 – 22. The fractions 1-8 presumably contained the proteins larger than 70 kDa. The concentration of the proteins was checked by SDS-PAGE (Figure 39).

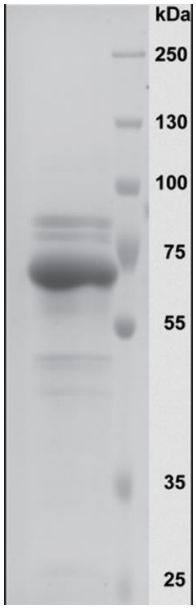

Figure 39: SDS-PAGE analysis of the CHOK1 conditioned medium. The largest band probably represents the leftover BSA (66 kDa) present in the medium. This analysis was performed in the Laboratory of Structural Biology (BIOCEV).

The inhibitory effect of individual fractions was determined. The major inhibitory effect was observed in the fractions 7, 8, and 9 (Figure 40). These fractions were mixed and submitted for the mass spectrometry analysis to identify the spectrum of proteins present.

Due to the fact that mass spectrometry analysis does not generally provide data for organisms without the existing protein database, the obtained amino acid sequences were first used in blast search against a database consisting of retroviral genes and CHOK1 proteome. These searches only yielded results with a high background probably caused by the carryover serum proteins present in the analyzed medium. We proceeded with the search using a personalized database, generated from retroviral envelope genes predicted in the CHOK1 genome. Due to their repetitive nature, retroviral gene products are generally not annotated in commonly used proteins databases such as Uniprot.

Seven peptides were identified by analysis of the mass spectrometry data using the CHOK1 endogenous retrovirus database. Importantly, two of the peptides were scored as high confidence hits (NW_003617793.1_49834_49998 and

NW_003614069.1_559612_559788). Each of the identified proteins was present only on one peptide, which might be caused by the short sequences provided in the database.

Analysis of protein sequences of these two hits showed that they originate from gammaretrovirus envelopes related to, but not identical, to FeLV. This is consistent with the previous hypothesis that the secreted inhibitors are soluble fragments of envelope glycoproteins, produced from ERVs in the hamster genome (Miller & Miller, 1992). These proteins would have the ability to bind to and saturate the amphotropic receptor and cause the infection block. Our results enable us to test specific hamster endogenous retrovirus loci for the capability to block the amphotropic MLV infection. These experiments are currently in progress.

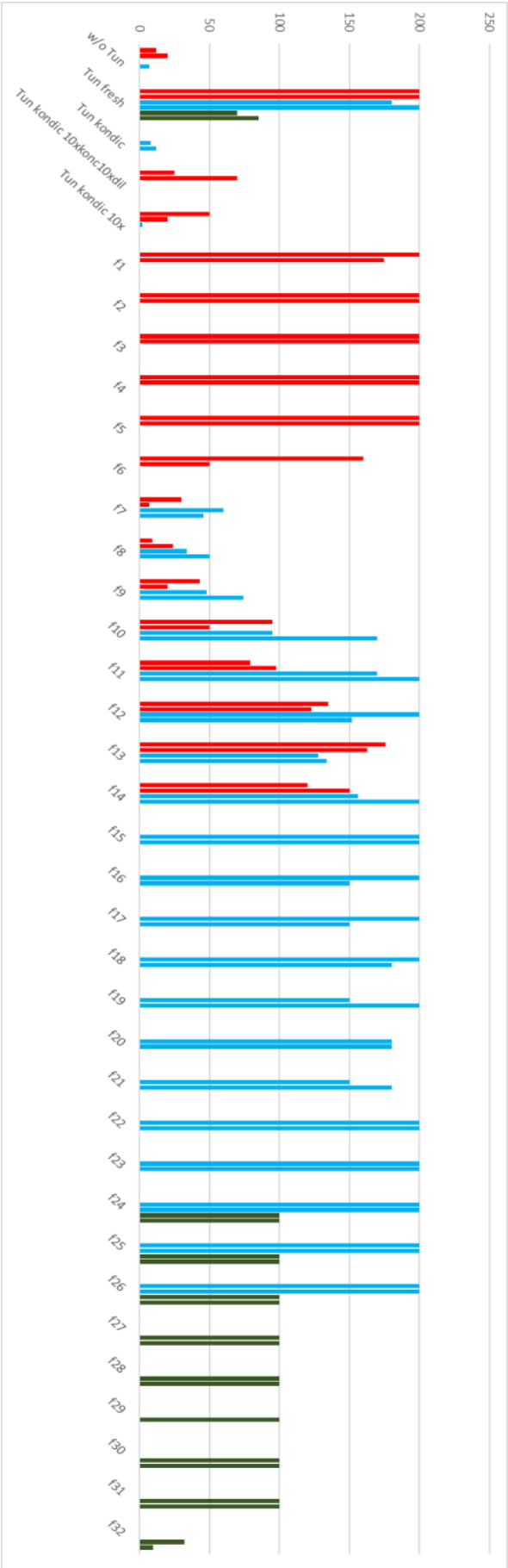

Figure 40 (left): Inhibitory effects of fractions of CHOK1 medium. The y axis depicts the number of GFP-positive foci. The assays were performed in duplicates. All of the infections were performed using the same virus with an amphotropic envelope. The fractions were tested in three consecutive experiments. The first experiment is depicted in red, second in light blue and third in dark green. In third experiment, half the amount of the virus was used compared to the amount used in testing the first two experiments. The first group of columns depicts the amount of foci if CHOK1 cells were infected without previous tunicamycin treatment. The second group of columns depicts the infectivity of the virus on the tunicamycin treated cells with fresh medium in the culture. The third group of columns depicts the infectivity of the of the virus on the tunicamycin treated cells with conditioned CHOK1 medium in the culture. The fourth and fifth group of columns depicts the inhibitory effect of the 10x concentrated CHOK1 conditioned medium; the 10x concentrated medium is further 10x diluted in the fourth column to check the strenght of the inhibition. All of the subsequent groups of columns depict the the infectivity of the virus on the tunicamycin treated cells with particular medium fractions in the culture. The GFP-positive foci were always counted three days post infection.

5 CONCLUSION

5.1 Screening for endogenous retrovirus presence

We performed several computational screens in order to find unusual endogenous retroviruses in vertebrate genomes. The most compelling cases from these screens were further characterized computationally and genetically.

5.1.1 ELVgv

We detected a novel endogenous Lentivirus (ELVgv) in the genome of *Galeopterus variegatus*, a close relative of primates. The outstanding evolutionary age of this novel lentiviral lineage (up to 60 MYA) was confirmed with three approaches:

- Orthologous integrations in extant dermopteran species were present (we detected an orthologous endogenous Lentivirus in the only other extant species – *Cynocephalus volans* – in the Dermoptera clade.).
- Sequence comparison and phylogenetic analysis of three complete ELVgv proviruses indicated the estimated age.
- LTR aging method confirmed the predicted age.

Considering this and the fact that endogenous lentiviruses are still rare (only four documented cases in mammals up to date), this finding provides new insight on the earliest lentiviral evolution and endogenization.

5.1.2 MINERVa

We detected remnants of endogenous Deltaretrovirus (MINERVa) in the genome of *Miniopterus natalensis*. This is the first report of an endogenous Deltaretrovirus. In addition, it is the first Deltaretrovirus described in the order Chiroptera. We detected and sequenced the provirus orthologues in other species of miniopterid bats.

- The MINERVa provirus is present in the Miniopterus genome only in one (diploid) copy.
- Orthologous MINERVa proviruses were detected in all other miniopterid species analyzed and not in species from other bat families.
- The described provirus contains a large internal deletion, encompassing the *pol* and *env* genes; however both LTRs are present in the sequence.

- Additional ORFs presumably coding for virus accessory genes were detected.
- The presence of presumed *rex* accessory gene was further indicated by predicting the characteristic RNA secondary structure of Rex Response element in the LTR of MINERVa.
- The predicted protein product of ORF present in the *gag* gene was not detected in the analyzed muscle sample by mass spectrometry.

5.2 Host-virus relationship of currently endogenizing CrERV

We were able to induce the endogenous retrovirus by co-cultivation of mule deer cells with susceptible human cells. We sequenced the induced virus and performed phylogenetic and sequence analyzes which confirmed that it is closely related to the young endogenous CrERV copies present in the deer genome. The induced virus shows xenotropic behavior despite the fact that CrERV is described as very effective in creating new endogenous copies.

- The induced CrERV particles sediment in the density gradient regions characteristic for retroviruses.
- The sequence of the induced provirus clusters closely with young ERVs in the mule deer genome.
- Retrovirus pseudotypes and marker rescue experiments indicate the presence of a block at the level of receptor-mediated virus entry, which could at least partly explain the xenotropic nature of CrERV. However, continuous struggle with low virus titers is preventing definitive conclusions to be made.
- Mule deer, fallow deer, and red deer primary cells all interact with CrERV in the similar fashion.
- Marker rescue experiments indicate the presence of the rescued provirus (GFP marker) several days after infection, but PERT assay on these cells indicate that the virus is probably inable to propagate in the deer cells.
- Nucleofection experiments indicate that the block of CrERV infection might also occur at the later stages of the infection.

- We tried to assemble several retroviral restriction factors from the mule deer genome, but currently we were successful only with SAMHD1 because we did not possess enough sequence data to assemble other restriction factors reliably.

5.3 Determining the infection block in the CHOK1 cells

We were successful in replicating the experiments, which lead to conclusion that a secreted protein factor produced by CHOK1 cells has an inhibitory effect on infection by a virus with amphotropic envelope glycoprotein. The preliminary analyzes of the medium fraction with the highest inhibitory activity are in agreement with the previously published proposal that the infection block is mediated by the presence of an endogenous retrovirus-encoded Env protein. We propose a competitive inhibition mediated by the defective product of the *env* gene of the presumably present endogenous retrovirus in the CHOK1 cells genome.

- We replicated the previously reported experiments rendering the susceptibility of CHOK1 cells to the infection by amphotropic retrovirus.
- We fractionated the proteins present in the medium conditioned by CHOK1 cells. The fractions of the conditioned CHOK1 medium containing inhibitory proteins with the size predicted in previous studies were tested and proved to be inhibitory in infections with amphotropic MLV virus.
- The mass spectrometry analysis of the fractions inhibiting the infection yielded data indicative of the fact that the inhibitory effect is mediated by the presence of an endogenous gammaretroviral *env* gene producing Env protein. The identification of the causative ERV element is the subject of our ongoing work.

6 SIGNIFICANCE OF RESULTS AND FUTURE PROSPECTS

We discovered the fourth endogenous Lentivirus lineage known up to date, present in the genome of *Galeopterus variegatus*. We showed that this virus lineage constitutes currently the oldest known Lentivirus. Based on ancient nature of the virus, structural studies of any of the proteins possibly produced by the ELVgv provirus might provide an insight into the evolution of viral genes and their products.

We detected remnants of endogenous Deltaretrovirus in the genome of *Miniopterus natalensis*. Even though the remnants of a provirus were found only in a single copy, its orthologues were found among Miniopteridae bats. Deltaretroviruses were the last genus of retroviruses lacking in an endogenous form across the sequenced genomes, so this finding is filling in the gap of evolutionary evidence.

We induced an endogenous mammalian gammaretrovirus (CrERV), which is presumed to be currently endogenizing in the genome of mule deer. This was the first study which aimed to analyze CrERV not only from sequential, but also virological perspective. We were partially successful in analyzing its xenotropic behavior. However, the results need to be further validated and the behavior presumably compared to other well characterized endogenous gammaretroviruses (e.g. PERV). Establishing CrERV as a model of mammalian ERV endogenization will enable us to characterize epigenetic modifications across a large spectrum of endogenous copies with various evolutionary ages (T. Hron et al, manuscript in preparation).

We detected presence of products of endogenous retroviruses in the culture medium from Chinese hamster ovary cells (CHOK1). These products presumably have an inhibitory effect on amphotropic retrovirus infection. Further causative studies (using CRISPR-Cas technology and ectopic overexpression the inhibitory product of the endogenous retrovirus) are needed in order to validate our current data and specify the inhibitory ERV elements in the CHOK1 genome.

It is indeniable that endogenous retroviruses shape the genetic information we bear within. For this, I believe that by studying the perplexity of the interactions between endogenous retroviruses and their hosts, we would be able to grasp the concept of our co-evolution. By understanding the concept and answering some of the questions from

the occurring pleiad, we would understand not only the evolution of viruses, but origin and development of the other forms of life as well.

7 INVOLVEMENT OF THE STUDENT IN THE PUBLICATIONS

The published papers and a currently reviewed manuscript associated with this thesis are in the Supplement of this thesis.

Publication 1: Endogenous Lentivirus in Malayan colugo (Galeopterus variegatus), a close relative of primates
Shared first authorship

- Help with the preparation of BLAST database
- Phylogeny analyzes

Publication 2: Life History of the Oldest Lentivirus: Characterization of ELVgv Integrations in the Dermopteran Genome
Second author

- PCR amplification and Sanger sequencing of the provirus sequences
- PCR amplification and Sanger sequencing of the control loci
- Participation in phylogenetic analyzes

Publication 3: Discovery of the first endogenous Deltaretrovirus, in the genome of long-fingered bats (Miniopteridae)
Shared first authorship

- PCR amplification and Sanger sequencing of the provirus sequences
- PCR amplification and Sanger sequencing of the control loci
- Annotation of the sequence
- Prediction of the secondary structures in LTRs

Publication 4: Induction and characterization of a replication competent cervid endogenous Gammaretrovirus (CrERV) from mule deer cells
Shared first authorship

- Maintenance of the co-culture, other cell cultures and virological assays
- Preparation of the cells for electron microscopy
- PERT assays
- Centrifugation methods
- Western blotting
- Participation in PCR amplifications and Sanger sequencing

7.1.1 Prospective publications

Publication 5: Characterization of xenotropic behavior of CrERV and its later replication blocks

- Pseudotyping experiments
- Marker rescue experiments

Publication 6: Elucidation of the block to amphotropic MLV infection in CHOK1

- Infection assays, testing of conditioned media fractions

8 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(3): 489-494.

Abascal, F., Zardoya, R., & Posada, D. 2005. ProtTest: selection of best-fit models of protein evolution. ***Bioinformatics***, 21(9): 2104-2105.

Abe, M., Fukuma, A., Yoshikawa, R., Miyazawa, T., & Yasuda, J. 2014. Inhibition of budding/release of porcine endogenous retrovirus. ***Microbiol Immunol***, 58(8): 432-438.

Ahn, J. 2016. Functional organization of human SAMHD1 and mechanisms of HIV-1 restriction. ***Biol Chem***, 397(4): 373-379.

Aiewsakun, P., & Katzourakis, A. 2015. Endogenous viruses: Connecting recent and ancient viral evolution. ***Virology***, 479-480: 26-37.

Aiewsakun, P., & Katzourakis, A. 2017. Marine origin of retroviruses in the early Palaeozoic Era. ***Nat Commun***, 8: 13954.

Albritton, L. M., Tseng, L., Scadden, D., & Cunningham, J. M. 1989. A putative murine ecotropic retrovirus receptor gene encodes a multiple membrane-spanning protein and confers susceptibility to virus infection. ***Cell***, 57(4): 659-666.

Alfano, N., Michaux, J., Morand, S., Aplin, K., Tsangaras, K., Lober, U., Fabre, P. H., Fitriana, Y., Semiadi, G., Ishida, Y., Helgen, K. M., Roca, A. L., Eiden, M. V., & Greenwood, A. D. 2016. Endogenous Gibbon Ape Leukemia Virus Identified in a Rodent (Melomys burtoni subsp.) from Wallacea (Indonesia). ***J Virol***, 90(18): 8169-8180.

Anderson, M. M., Lauring, A. S., Burns, C. C., & Overbaugh, J. 2000. Identification of a cellular cofactor required for infection by feline leukemia virus. ***Science***, 287(5459): 1828-1830.

Armezzani, A., Arnaud, F., Caporale, M., di Meo, G., Iannuzzi, L., Murgia, C., & Palmarini, M. 2011. The signal peptide of a recently integrated endogenous sheep betaretrovirus envelope plays a major role in eluding gag-mediated late restriction. ***J Virol***, 85(14): 7118-7128.

Arnaud, F., Black, S. G., Murphy, L., Griffiths, D. J., Neil, S. J., Spencer, T. E., & Palmarini, M. 2010. Interplay between ovine bone marrow stromal cell antigen 2/tetherin and endogenous retroviruses. ***J Virol***, 84(9): 4415-4425.

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(11): e170.

Arnaud, F., Murcia, P. R., & Palmarini, M. 2007b. Mechanisms of late restriction induced by an endogenous retrovirus. ***J Virol***, 81(20): 11441-11451.

Aswad, A., & Katzourakis, A. 2014. The first endogenous herpesvirus, identified in the tarsier genome, and novel sequences from primate rhadinoviruses and lymphocryptoviruses. ***PLoS Genet***, 10(6): e1004332.

Avila-Arcos, M. C., Ho, S. Y., Ishida, Y., Nikolaidis, N., Tsangaras, K., Honig, K., Medina, R., Rasmussen, M., Fordyce, S. L., Calvignac-Spencer, S., Willerslev, E., Gilbert, M. T., Helgen, K. M., Roca, A. L., & Greenwood, A. D. 2013. One hundred twenty years of koala retrovirus evolution determined from museum skins. ***Mol Biol Evol***, 30(2): 299-304.

Bae, E. H., & Jung, Y. T. 2014a. Comparison of the effects of retroviral restriction factors involved in resistance to porcine endogenous retrovirus. ***J Microbiol Biotechnol***, 24(4): 577-583.

Bae, E. H., & Jung, Y. T. 2014b. Tetherins of various species inhibit the release of porcine endogenous retrovirus from human cells. ***Acta Virol***, 58(1): 53-60.

Baltimore, D. 1995. Discovery of the reverse transcriptase. ***FASEB J***, 9(15): 1660-1663.

Ban, J., Portetelle, D., Altaner, C., Horion, B., Milan, D., Krchnak, V., Burny, A., & Kettmann, R. 1993. Isolation and characterization of a 2.3-kilobase-pair cDNA fragment encoding the binding domain of the bovine leukemia virus cell receptor. ***J Virol***, 67(2): 1050-1057.

Banerji, N., Kapur, V., & Kanjilal, S. 2007. Association of germ-line polymorphisms in the feline p53 gene with genetic predisposition to vaccine-associated feline sarcoma. ***J Hered***, 98(5): 421-427.

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(1): 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(1): 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(24): 13880-13890.

Baycin-Hizal, D., Tabb, D. L., Chaerkady, R., Chen, L., Lewis, N. E., Nagarajan, H., Sarkaria, V., Kumar, A., Wolozny, D., Colao, J., Jacobson, E., Tian, Y., O'Meally, R. N., Krag, S. S., Cole, R. N., Palsson, B. O., Zhang, H., & Betenbaugh, M. 2012. Proteomic analysis of Chinese hamster ovary cells. ***J Proteome Res***, 11(11): 5265-5276.

Bednarik, D. P., Cook, J. A., & Pitha, P. M. 1990. Inactivation of the HIV LTR by DNA CpG methylation: evidence for a role in latency. ***EMBO J***, 9(4): 1157-1164.

Bibollet-Ruche, F., Bailes, E., Gao, F., Pourrut, X., Barlow, K. L., Clewley, J. P., Mwenda, J. M., Langat, D. K., Chege, G. K., McClure, H. M., Mpoudi-Ngole, E., Delaporte, E., Peeters, M., Shaw, G. M., Sharp, P. M., & Hahn, B. H. 2004. New simian immunodeficiency virus infecting De Brazza's monkeys (Cercopithecus neglectus): evidence for a cercopithecus monkey virus clade. ***J Virol***, 78(14): 7748-7762.

Biek, R., Drummond, A. J., & Poss, M. 2006. A virus reveals population structure and recent demographic history of its carnivore host. ***Science***, 311(5760): 538-541.

Bittner, J. J. 1936. Some Possible Effects of Nursing on the Mammary Gland Tumor Incidence in Mice. ***Science***, 84(2172): 162.

Blazkova, J., Trejbalova, K., Gondois-Rey, F., Halfon, P., Philibert, P., Guiguen, A., Verdin, E., Olive, D., Van Lint, C., Hejnar, J., & Hirsch, I. 2009. CpG methylation controls reactivation of HIV from latency. ***PLoS Pathog***, 5(8): e1000554.

Blissenbach, M., Grewe, B., Hoffmann, B., Brandt, S., & Uberla, K. 2010. Nuclear RNA export and packaging functions of HIV-1 Rev revisited. ***J Virol***, 84(13): 6598-6604.

Boid, R., McOrist, S., Jones, T. W., Easterbee, N., Hubbard, A. L., & Jarrett, O. 1991. Isolation of FeLV from a wild felid (Felis silvestris). ***Vet Rec***, 128(11): 256.

Bowser, P. R., Wolfe, M. J., Forney, J. L., & Wooster, G. A. 1988. Seasonal prevalence of skin tumors from walleye (Stizostedion vitreum) from Oneida Lake, New York. ***J Wildl Dis***, 24(2): 292-298.

Bulli, L., Apolonia, L., Kutzner, J., Pollpeter, D., Goujon, C., Herold, N., Schwarz, S. M., Giernat, Y., Keppler, O. T., Malim, M. H., & Schaller, T. 2016. Complex Interplay between HIV-1 Capsid and MX2-Independent Alpha Interferon-Induced Antiviral Factors. ***J Virol***, 90(16): 7469-7480.

Bush, S., & Tebit, D. M. 2015. HIV-1 Group O Origin, Evolution, Pathogenesis, and Treatment: Unraveling the Complexity of an Outlier 25 Years Later. ***AIDS Rev***, 17(3): 147-158.

Campbell, E. M., Weingart, J., Sette, P., Opp, S., Sastri, J., O'Connor, S. K., Talley, S., Diaz-Griffero, F., Hirsch, V., & Bouamr, F. 2015. TRIM5alpha-Mediated Ubiquitin Chain Conjugation Is Required for Inhibition of HIV-1 Reverse Transcription and Capsid Destabilization. ***J Virol***, 90(4): 1849-1857.

Canfield, P. J., Sabine, J. M., & Love, D. N. 1988. Virus particles associated with leukaemia in a koala. ***Aust Vet J***, 65(10): 327-328.

Cohen, C. J., Lock, W. M., & Mager, D. L. 2009. Endogenous retroviral LTRs as promoters for human genes: a critical assessment. ***Gene***, 448(2): 105-114.

Cohen, S., Au, S., & Pante, N. 2011. How viruses access the nucleus. ***Biochim Biophys Acta***, 1813(9): 1634-1645.

Conticello, S. G., Thomas, C. J., Petersen-Mahrt, S. K., & Neuberger, M. S. 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. ***Mol Biol Evol***, 22(2): 367-377.

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(1): 262-276.

Cornelis, G., Vernochet, C., Carradec, Q., Souquere, S., Mulot, B., Catzeflis, F., Nilsson, M. A., Menzies, B. R., Renfree, M. B., Pierron, G., Zeller, U., Heidmann, O., Dupressoir, A., & Heidmann, T. 2015. Retroviral envelope gene captures and syncytin exaptation for placentation in marsupials. ***Proc Natl Acad Sci U S A***, 112(5): E487-496.

Cui, J., & Holmes, E. C. 2012a. Endogenous lentiviruses in the ferret genome. ***J Virol***, 86(6): 3383-3385.

Cui, J., & Holmes, E. C. 2012b. Evidence for an endogenous papillomavirus-like element in the platypus genome. ***J Gen Virol***, 93(Pt 6): 1362-1366.

Cui, J., Tachedjian, G., Tachedjian, M., Holmes, E. C., Zhang, S., & Wang, L. F. 2012. Identification of diverse groups of endogenous gammaretroviruses in mega- and microbats. ***J Gen Virol***, 93(Pt 9): 2037-2045.

Cui, J., & Wang, L. F. 2015. Genomic Mining Reveals Deep Evolutionary Relationships between Bornaviruses and Bats. ***Viruses***, 7(11): 5792-5800.

Cui, P., Lober, U., Alquezar-Planas, D. E., Ishida, Y., Courtiol, A., Timms, P., Johnson, R. N., Lenz, D., Helgen, K. M., Roca, A. L., Hartman, S., & Greenwood, A. D. 2016. Comprehensive profiling of retroviral integration sites using target enrichment methods from historical koala samples without an assembled reference genome. ***PeerJ***, 4: e1847.

Cunningham, M. W., Brown, M. A., Shindle, D. B., Terrell, S. P., Hayes, K. A., Ferree, B. C., McBride, R. T., Blankenship, E. L., Jansen, D., Citino, S. B., Roelke, M. E., Kiltie, R. A., Troyer, J. L., & O'Brien, S. J. 2008. Epizootiology and management of feline leukemia virus in the Florida puma. ***J Wildl Dis***, 44(3): 537-552.

Czernilofsky, A. P., Levinson, A. D., Varmus, H. E., Bishop, J. M., Tischer, E., & Goodman, H. M. 1980. Nucleotide sequence of an avian sarcoma virus oncogene (src) and proposed amino acid sequence for gene product. ***Nature***, 287(5779): 198-203.

Dalgleish, A. G., Beverley, P. C., Clapham, P. R., Crawford, D. H., Greaves, M. F., & Weiss, R. A. 1984. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. ***Nature***, 312(5996): 763-767.

Davis, H. E., Morgan, J. R., & Yarmush, M. L. 2002. Polybrene increases retrovirus gene transfer efficiency by enhancing receptor-independent virus adsorption on target cell membranes. ***Biophys Chem***, 97(2-3): 159-172.

Delport, W., Poon, A. F., Frost, S. D., & Kosakovsky Pond, S. L. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. ***Bioinformatics***, 26(19): 2455-2457.

Denner, J., & Young, P. R. 2013. Koala retroviruses: characterization and impact on the life of koalas. ***Retrovirology***, 10: 108.

Derse, D., Crise, B., Li, Y., Princler, G., Lum, N., Stewart, C., McGrath, C. F., Hughes, S. H., Munroe, D. J., & Wu, X. 2007. Human T-cell leukemia virus type 1 integration target sites in the human genome: comparison with those of other retroviruses. ***J Virol***, 81(12): 6731-6741.

Dimmic, M. W., Rest, J. S., Mindell, D. P., & Goldstein, R. A. 2002. rtREV: an amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. ***J Mol Evol***, 55(1): 65-73.

Doehle, B. P., Schafer, A., Wiegand, H. L., Bogerd, H. P., & Cullen, B. R. 2005. Differential sensitivity of murine leukemia virus to APOBEC3-mediated inhibition is governed by virion exclusion. ***J Virol***, 79(13): 8201-8207.

Donahue, P. R., Quackenbush, S. L., Gallo, M. V., deNoronha, C. M., Overbaugh, J., Hoover, E. A., & Mullins, J. I. 1991. Viral genetic determinants of T-cell killing and immunodeficiency disease induction by the feline leukemia virus FeLV-FAIDS. ***J Virol***, 65(8): 4461-4469.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. ***Nucleic Acids Res***, 32(5): 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(5): 2787-2796.

Elleder, D., Pavlicek, A., Paces, J., & Hejnar, J. 2002. Preferential integration of human immunodeficiency virus type 1 into genes, cytogenetic R bands and GC-rich DNA regions: insight from the human genome sequence. ***FEBS Lett***, 517(1-3): 285-286.

Elleder, D., Plachy, J., Hejnar, J., Geryk, J., & Svoboda, J. 2004. Close linkage of genes encoding receptors for subgroups A and C of avian sarcoma/leucosis virus on chicken chromosome 28. ***Anim Genet***, 35(3): 176-181.

Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A. J., Heidmann, T., & Schwartz, O. 2005. APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. ***Nature***, 433(7024): 430-433.

Fabryova, H., Hron, T., Kabickova, H., Poss, M., & Elleder, D. 2015. Induction and characterization of a replication competent cervid endogenous gammaretrovirus (CrERV) from mule deer cells. ***Virology***, 485: 96-103.

Fadel, H. J., Saenz, D. T., Guevara, R., von Messling, V., Peretz, M., & Poeschla, E. M. 2012. Productive replication and evolution of HIV-1 in ferret cells. ***J Virol***, 86(4): 2312-2322.

Farnet, C. M., & Haseltine, W. A. 1990. Integration of human immunodeficiency virus type 1 DNA in vitro. ***Proc Natl Acad Sci U S A***, 87(11): 4164-4168.

Feschotte, C., & Gilbert, C. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. ***Nat Rev Genet***, 13(4): 283-296.

Frankel, W. N., Stoye, J. P., Taylor, B. A., & Coffin, J. M. 1990. A linkage map of endogenous murine leukemia proviruses. ***Genetics***, 124(2): 221-236.

Genome, K. C. o. S. 2009. Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. ***J Hered***, 100(6): 659-674.

Genomes Project, C., Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., Gibbs, R. A., Hurles, M. E., & McVean, G. A. 2010. A map of human genome variation from population-scale sequencing. ***Nature***, 467(7319): 1061-1073.

Gifford, R. J., Katzourakis, A., Tristem, M., Pybus, O. G., Winters, M., & Shafer, R. W. 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(51): 20362-20367.

Gilbert, C., Maxfield, D. G., Goodman, S. M., & Feschotte, C. 2009. Parallel germline infiltration of a lentivirus in two Malagasy lemurs. ***PLoS Genet***, 5(3): e1000425.

Gilbert, C., Meik, J. M., Dashevsky, D., Card, D. C., Castoe, T. A., & Schaack, S. 2014. Endogenous hepadnaviruses, bornaviruses and circoviruses in snakes. ***Proc Biol Sci***, 281(1791): 20141122.

Gilbert, J. M., Bates, P., Varmus, H. E., & White, J. M. 1994. The receptor for the subgroup A avian leukosis-sarcoma viruses binds to subgroup A but not to subgroup C envelope glycoprotein. ***J Virol***, 68(9): 5623-5628.

Goldstone, D. C., Yap, M. W., Robertson, L. E., Haire, L. F., Taylor, W. R., Katzourakis, A., Stoye, J. P., & Taylor, I. A. 2010. Structural and functional analysis of prehistoric lentiviruses uncovers an ancient molecular interface. ***Cell Host Microbe***, 8(3): 248-259.

Goodrich, D. W., & Duesberg, P. H. 1990. Retroviral recombination during reverse transcription. ***Proc Natl Acad Sci U S A***, 87(6): 2052-2056.

Goto, T., Kennel, S. J., Abe, M., Takishita, M., Kosaka, M., Solomon, A., & Saito, S. 1994. A novel membrane antigen selectively expressed on terminally differentiated human B cells. ***Blood***, 84(6): 1922-1930.

Goujon, C., Moncorge, O., Bauby, H., Doyle, T., Ward, C. C., Schaller, T., Hue, S., Barclay, W. S., Schulz, R., & Malim, M. H. 2013. Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. ***Nature***, 502(7472): 559-562.

Grewe, B., & Uberla, K. 2010. The human immunodeficiency virus type 1 Rev protein: menage a trois during the early phase of the lentiviral replication cycle. ***J Gen Virol***, 91(Pt 8): 1893-1897.

Guo, F., Cen, S., Niu, M., Saadatmand, J., & Kleiman, L. 2006. Inhibition of tRNA(3)(Lys)-primed reverse transcription by human APOBEC3G during human immunodeficiency virus type 1 replication. ***J Virol***, 80(23): 11710-11722.

Han, G. Z., & Worobey, M. 2012. Endogenous lentiviral elements in the weasel family (Mustelidae). ***Mol Biol Evol***, 29(10): 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(1): 211-215.

Hanger, J. J., Bromham, L. D., McKee, J. J., O'Brien, T. M., & Robinson, W. F. 2000. The nucleotide sequence of koala (Phascolarctos cinereus) retrovirus: a novel type C endogenous virus related to Gibbon ape leukemia virus. ***J Virol***, 74(9): 4264-4272.

Hasan, M., & Yan, N. 2014. Safeguard against DNA sensing: the role of TREX1 in HIV-1 infection and autoimmune diseases. ***Frontiers in Microbiology***, 5.

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.

Hefzi, H., Ang, K. S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C. A., Baycin-Hizal, D., Huang, Y., Ley, D., Martinez, V. S., Kyriakopoulos, S., Jimenez, N. E., Zielinski, D. C., Quek, L. E., Wulff, T., Arnsdorf, J., Li, S., Lee, J. S., Paglia, G., Loira, N., Spahn, P. N., Pedersen, L. E., Gutierrez, J. M., King, Z. A., Lund, A. M., Nagarajan, H., Thomas, A., Abdel-Haleem, A. M., Zanghellini, J., Kildegaard, H. F., Voldborg, B. G., Gerdtzen, Z. P., Betenbaugh, M. J., Palsson, B. O., Andersen, M. R., Nielsen, L. K., Borth, N., Lee, D. Y., & Lewis, N. E. 2016. A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. ***Cell Syst***, 3(5): 434-443 e438.

Henzy, J. E., Gifford, R. J., Kenaley, C. P., & Johnson, W. E. 2016. An Intact Retroviral Gene Conserved in Spiny-Rayed Fishes for over 100 My. ***Mol Biol Evol***.

Herrmann, A., Happel, A. U., & Gramberg, T. 2016. SAMHD1 in Retroviral Restriction and Innate Immune Sensing--Should We Leash the Hound? ***Curr HIV Res***, 14(3): 225-234.

Hizi, A., & Herzig, E. 2015. dUTPase: the frequently overlooked enzyme encoded by many retroviruses. ***Retrovirology***, 12: 70.

Holmes, R. K., Koning, F. A., Bishop, K. N., & Malim, M. H. 2007. APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation. Comparisons with APOBEC3G. ***J Biol Chem***, 282(4): 2587-2595.

Holt, M. P., Shevach, E. M., & Punkosdy, G. A. 2013. Endogenous mouse mammary tumor viruses (mtv): new roles for an old virus in cancer, infection, and immunity. ***Front Oncol***, 3: 287.

Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P., Gojobori, T., Coffin, J. M., & Tomonaga, K. 2010. Endogenous non-retroviral RNA virus elements in mammalian genomes. ***Nature***, 463(7277): 84-87.

Hoss, M., Robins, P., Naven, T. J. P., Pappin, D. J. C., Sgouros, J., & Lindahl, T. 1999. A human DNA editing enzyme homologous to the Escherichia coli DnaQ/MutD protein. ***Embo Journal***, 18(13): 3868-3875.

Hron, T., Fabryova, H., Paces, J., & Elleder, D. 2014. Endogenous lentivirus in Malayan colugo (Galeopterus variegatus), a close relative of primates. ***Retrovirology***, 11: 84.

Huder, J. B., Boni, J., Hatt, J. M., Soldati, G., Lutz, H., & Schupbach, J. 2002. Identification and characterization of two closely related unclassifiable endogenous retroviruses in pythons (Python molurus and Python curtus). ***J Virol***, 76(15): 7607-7615.

Huelsenbeck, J. P., & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. ***Bioinformatics***, 17(8): 754-755.

Hussain, A., Wesley, C., Khalid, M., Chaudhry, A., & Jameel, S. 2008. Human immunodeficiency virus type 1 Vpu protein interacts with CD74 and modulates major histocompatibility complex class II presentation. ***J Virol***, 82(2): 893-902.

Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., Kao, R. R., Pemberton, J. M., Beraldi, D., Stear, M. J., Alberti, A., Pittau, M., Iannuzzi, L., Banabazi, M. H., Kazwala, R. R., Zhang, Y. P., Arranz, J. J., Ali, B. A., Wang, Z., Uzun, M., Dione, M. M., Olsaker, I., Holm, L. E., Saarma, U., Ahmad, S., Marzanov, N., Eythorsdottir, E., Holland, M. J., Ajmone-Marsan, P., Bruford, M. W., Kantanen, J., Spencer, T. E., & Palmarini, M. 2009. Revealing the history of sheep domestication using retrovirus integrations. ***Science***, 324(5926): 532-536.

Chuong, E. B., Elde, N. C., & Feschotte, C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. ***Science***, 351(6277): 1083-1087.

Imakawa, K., Nakagawa, S., & Miyazawa, T. 2015. Baton pass hypothesis: successive incorporation of unconserved endogenous retroviral genes for placentation during mammalian evolution. ***Genes Cells***, 20(10): 771-788.

Ishida, Y., Zhao, K., Greenwood, A. D., & Roca, A. L. 2015. Proliferation of endogenous retroviruses in the early stages of a host germ line invasion. ***Mol Biol Evol***, 32(1): 109-120.

Ito, J., Baba, T., Kawasaki, J., & Nishigaki, K. 2015. Ancestral Mutations Acquired in Refrex-1, a Restriction Factor against Feline Retroviruses, during its Cooption and Domestication. ***J Virol***, 90(3): 1470-1485.

Ito, J., Watanabe, S., Hiratsuka, T., Kuse, K., Odahara, Y., Ochi, H., Kawamura, M., & Nishigaki, K. 2013. Refrex-1, a soluble restriction factor against feline endogenous and exogenous retroviruses. ***J Virol***, 87(22): 12029-12040.

Janecka, J. E., Helgen, K. M., Lim, N. T. L., Baba, M., Izawa, M., & Murphy, W. J. 2008. Evidence for multiple species of Sunda colugo. ***Current Biology***, 18(21): R1001-R1002.

Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J., & Navaratnam, N. 2002. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. ***Genomics***, 79(3): 285-296.

Jha, A. R., Pillai, S. K., York, V. A., Sharp, E. R., Storm, E. C., Wachter, D. J., Martin, J. N., Deeks, S. G., Rosenberg, M. G., Nixon, D. F., & Garrison, K. E. 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(11): 2617-2626.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. 2008. NCBI BLAST: a better web interface. ***Nucleic Acids Res***, 36(Web Server issue): W5-9.

Johnson, W. E. 2015. Endogenous Retroviruses in the Genomics Era. ***Annu Rev Virol***, 2(1): 135-159.

Johnson, W. E., & Coffin, J. M. 1999. Constructing primate phylogenies from ancient retrovirus sequences. ***Proc Natl Acad Sci U S A***, 96(18): 10254-10260.

Jolicoeur, P., & Baltimore, D. 1976. Effect of Fv-1 gene product on proviral DNA formation and integration in cells infected with murine leukemia viruses. ***Proc Natl Acad Sci U S A***, 73(7): 2236-2240.

Jung, Y. T., Lyu, M. S., Buckler-White, A., & Kozak, C. A. 2002. Characterization of a polytropic murine leukemia virus proviral sequence associated with the virus resistance gene Rmcf of DBA/2 mice. ***J Virol***, 76(16): 8218-8224.

Kalter, S. S., Helmke, R. J., Heberling, R. L., Panigel, M., Fowler, A. K., Strickland, J. E., & Hellman, A. 1973. Brief communication: C-type particles in normal human placentas. ***J Natl Cancer Inst***, 50(4): 1081-1084.

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(2): 173-187.

Kambol, R., Kabat, P., & Tristem, M. 2003. Complete nucleotide sequence of an endogenous retrovirus from the amphibian, Xenopus laevis. ***Virology***, 311(1): 1-6.

Kane, M., Yadav, S. S., Bitzegeio, J., Kutluay, S. B., Zang, T., Wilson, S. J., Schoggins, J. W., Rice, C. M., Yamashita, M., Hatziioannou, T., & Bieniasz, P. D. 2013. MX2 is an interferon-induced inhibitor of HIV-1 infection. ***Nature***, 502(7472): 563-566.

Kang, J., Ido, E., Pawling, J., Beutner, U., Huber, B. T., & Hozumi, N. 1994. Expression of Mtv-7 sag gene in vivo using a retroviral vector results in selective inactivation of superantigen reactive T cells. ***J Immunol***, 152(3): 1039-1046.

Katzourakis, A., & Gifford, R. J. 2010. Endogenous viral elements in animal genomes. ***PLoS Genet***, 6(11): e1001191.

Katzourakis, A., Gifford, R. J., Tristem, M., Gilbert, M. T., & Pybus, O. G. 2009. Macroevolution of complex retroviruses. ***Science***, 325(5947): 1512.

Katzourakis, A., Rambaut, A., & Pybus, O. G. 2005. The evolutionary dynamics of endogenous retroviruses. ***Trends Microbiol***, 13(10): 463-468.

Katzourakis, A., Tristem, M., Pybus, O. G., & Gifford, R. J. 2007. Discovery and analysis of the first endogenous lentivirus. ***Proc Natl Acad Sci U S A***, 104(15): 6261-6265.

Kawakami, T. G., Kollias, G. V., Jr., & Holmberg, C. 1980. Oncogenicity of gibbon type-C myelogenous leukemia virus. ***Int J Cancer***, 25(5): 641-646.

Keckesova, Z., Ylinen, L. M., Towers, G. J., Gifford, R. J., & 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(1): 7-11.

Keeble, A. H., Khan, Z., Forster, A., & James, L. C. 2008. TRIM21 is an IgG receptor that is structurally, thermodynamically, and kinetically conserved. ***Proc Natl Acad Sci U S A***, 105(16): 6045-6050.

Kimsa, M. C., Strzalka-Mrozik, B., Kimsa, M. W., Gola, J., Nicholson, P., Lopata, K., & Mazurek, U. 2014. Porcine endogenous retroviruses in xenotransplantation--molecular aspects. ***Viruses***, 6(5): 2062-2083.

Kinney, M. E., & Pye, G. W. 2016. Koala Retrovirus: A Review. ***J Zoo Wildl Med***, 47(2): 387-396.

Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J. C., & Montagnier, L. 1984. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. ***Nature***, 312(5996): 767-768.

Klucking, S., Adkins, H. B., & Young, J. A. 2002. Resistance to infection by subgroups B, D, and E avian sarcoma and leukosis viruses is explained by a premature stop codon within a resistance allele of the tvb receptor gene. ***J Virol***, 76(15): 7918-7921.

Klymiuk, N., Muller, M., Brem, G., & Aigner, B. 2003. Characterization of endogenous retroviruses in sheep. ***J Virol***, 77(20): 11268-11273.

Koepfli, K. P., Paten, B., Genome, K. C. o. S., & O'Brien, S. J. 2015. The Genome 10K Project: a way forward. ***Annu Rev Anim Biosci***, 3: 57-111.

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(1): 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(1): 345-351.

Kosakovsky Pond, S. L., Murrell, B., Fourment, M., Frost, S. D., Delport, W., & Scheffler, K. 2011. A random effects branch-site model for detecting episodic diversifying selection. ***Mol Biol Evol***, 28(11): 3033-3043.

Kozak, C. A. 2014. Origins of the endogenous and infectious laboratory mouse gammaretroviruses. ***Viruses***, 7(1): 1-26.

Kuwata, T., Kaori, T., Enomoto, I., Yoshimura, K., & Matsushita, S. 2013. Increased infectivity in human cells and resistance to antibody-mediated neutralization by truncation of the SIV gp41 cytoplasmic tail. ***Front Microbiol***, 4: 117.

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.

Laguette, N., Bregnard, C., Benichou, S., & Basmaciogullari, S. 2010. Human immunodeficiency virus (HIV) type-1, HIV-2 and simian immunodeficiency virus Nef proteins. ***Mol Aspects Med***, 31(5): 418-433.

LaPierre, L. A., Holzschu, D. L., Wooster, G. A., Bowser, P. R., & Casey, J. W. 1998. Two closely related but distinct retroviruses are associated with walleye discrete epidermal hyperplasia. ***J Virol***, 72(4): 3484-3490.

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(16): 8868-8877.

Le Bao, D. E., Raunaq Malhotra , Michael DeGiorgio, Theodora Maravegias, Lindsay Horvath, Laura Carrel, Colin Gillin, Tomáš Hron, Helena Fábryová, David R. Hunter, Mary Poss. 2014. Computational and Statistical Analyzes of Insertional Polymorphic Endogenous Retroviruses in a Non-Model Organism. ***Computation***, 2(4): 221-245.

Lemos de Matos, A., de Sousa-Pereira, P., Lissovsky, A. A., van der Loo, W., Melo-Ferreira, J., Cui, J., & Esteves, P. J. 2015. Endogenization of mouse mammary tumor virus (MMTV)-like elements in genomes of pikas (Ochotona sp.). ***Virus Res***, 210: 22-26.

Lerner, D. L., Wagaman, P. C., Phillips, T. R., Prospero-Garcia, O., Henriksen, S. J., Fox, H. S., Bloom, F. E., & Elder, J. H. 1995. Increased mutation frequency of feline immunodeficiency virus lacking functional deoxyuridine-triphosphatase. ***Proc Natl Acad Sci U S A***, 92(16): 7480-7484.

Lewis, N. E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., Bordbar, A., Roth, A. M., Rosenbloom, J., Bian, C., Xie, M., Chen, W., Li, N., Baycin-Hizal, D., Latif, H., Forster, J., Betenbaugh, M. J., Famili, I., Xu, X., Wang, J., & Palsson, B. O. 2013. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. ***Nat Biotechnol***, 31(8): 759-765.

Li, W., Yap, M. W., Voss, V., & Stoye, J. P. 2016. Expression levels of Fv1: effects on retroviral restriction specificities. ***Retrovirology***, 13(1): 42.

Liao, W., Hong, S. H., Chan, B. H., Rudolph, F. B., Clark, S. C., & Chan, L. 1999. APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. ***Biochem Biophys Res Commun***, 260(2): 398-404.

Lichtenstein, D. L., Rushlow, K. E., Cook, R. F., Raabe, M. L., Swardson, C. J., Kociba, G. J., Issel, C. J., & Montelaro, R. C. 1995. Replication in vitro and in vivo of an equine infectious anemia virus mutant deficient in dUTPase activity. ***J Virol***, 69(5): 2881-2888.

Liu, Z., Pan, Q., Ding, S., Qian, J., Xu, F., Zhou, J., Cen, S., Guo, F., & Liang, C. 2013. The interferon-inducible MxB protein inhibits HIV-1 infection. ***Cell Host Microbe***, 14(4): 398-410.

Lochelt, M., Romen, F., Bastone, P., Muckenfuss, H., Kirchner, N., Kim, Y. B., Truyen, U., Rosler, U., Battenberg, M., Saib, A., Flory, E., Cichutek, K., & Munk, C. 2005. The antiretroviral activity of APOBEC3 is inhibited by the foamy virus accessory Bet protein. ***Proc Natl Acad Sci U S A***, 102(22): 7982-7987.

Lounkova, A., Draberova, E., Senigl, F., Trejbalova, K., Geryk, J., Hejnar, J., & Svoboda, J. 2014. Molecular events accompanying rous sarcoma virus rescue from rodent cells and the role of viral gene complementation. ***J Virol***, 88(6): 3505-3515.

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(2): 185-200.

Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., Firth, A., Singer, O., Trono, D., & Pfaff, S. L. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. ***Nature***, 487(7405): 57-63.

Mackey, L., Jarrett, W., Jarrett, O., & Laird, H. 1975. Anemia associated with feline leukemia virus infection in cats. ***J Natl Cancer Inst***, 54(1): 209-217.

Maddon, P. J., Dalgleish, A. G., McDougal, J. S., Clapham, P. R., Weiss, R. A., & Axel, R. 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. ***Cell***, 47(3): 333-348.

Makowka, L., Wu, G. D., Hoffman, A., Podesta, L., Sher, L., Tuso, P. J., Breda, M., Chapman, F. A., Cosenza, C., Yasunaga, C., & et al. 1994. Immunohistopathologic lesions associated with the rejection of a pig-to-human liver xenograft. ***Transplant Proc***, 26(3): 1074-1075.

Malfavon-Borja, R., & Feschotte, C. 2015. Fighting fire with fire: endogenous retrovirus envelopes as restriction factors. ***J Virol***, 89(8): 4047-4050.

Malicorne, S., Vernochet, C., Cornelis, G., Mulot, B., Delsuc, F., Heidmann, O., Heidmann, T., & Dupressoir, A. 2016. Genome-Wide Screening of Retroviral Envelope Genes in the Nine-Banded Armadillo (Dasypus novemcinctus, Xenarthra) Reveals an Unfixed Chimeric Endogenous Betaretrovirus Using the ASCT2 Receptor. ***J Virol***, 90(18): 8132-8149.

Marino, D., Perkovic, M., Hain, A., Jaguva Vasudevan, A. A., Hofmann, H., Hanschmann, K. M., Muhlebach, M. D., Schumann, G. G., Konig, R., Cichutek, K., Haussinger, D., & Munk, C. 2016. APOBEC4 Enhances the Replication of HIV-1. ***PLoS One***, 11(6): e0155422.

Marker, L., Munson, L., Basson, P. A., & Quackenbush, S. 2003. Multicentric T-cell lymphoma associated with feline leukemia virus infection in a captive namibian cheetah (Acinonyx jubatus). ***J Wildl Dis***, 39(3): 690-695.

Martin, G. S. 2004. The road to Src. ***Oncogene***, 23(48): 7910-7917.

Mattiuzzo, G., Ivol, S., & Takeuchi, Y. 2010. Regulation of porcine endogenous retrovirus release by porcine and human tetherins. ***J Virol***, 84(5): 2618-2622.

Mattiuzzo, G., & Takeuchi, Y. 2010. Suboptimal porcine endogenous retrovirus infection in non-human primate cells: implication for preclinical xenotransplantation. ***PLoS One***, 5(10): e13203.

Mattiuzzo, G., Takeuchi, Y., & Scobie, L. 2012. Potential zoonotic infection of porcine endogenous retrovirus in xenotransplantation. ***Methods Mol Biol***, 885: 263-279.

Mbisa, J. L., Barr, R., Thomas, J. A., Vandegraaff, N., Dorweiler, I. J., Svarovskaia, E. S., Brown, W. L., Mansky, L. M., Gorelick, R. J., Harris, R. S., Engelman, A., & Pathak, V. K. 2007. Human immunodeficiency virus type 1 cDNAs produced in the presence of APOBEC3G exhibit defects in plus-strand DNA transfer and integration. ***J Virol***, 81(13): 7099-7110.

McNally, M. M., Wahlin, K. J., & Canto-Soler, M. V. 2010. Endogenous expression of ASLV viral proteins in specific pathogen free chicken embryos: relevance for the developmental biology research field. ***BMC Dev Biol***, 10: 106.

Meli, M. L., Cattori, V., Martinez, F., Lopez, G., Vargas, A., Simon, M. A., Zorrilla, I., Munoz, A., Palomares, F., Lopez-Bao, J. V., Pastor, J., Tandon, R., Willi, B., Hofmann-Lehmann, R., & Lutz, H. 2009. Feline leukemia virus and other pathogens as important threats to the survival of the critically endangered Iberian lynx (Lynx pardinus). ***PLoS One***, 4(3): e4744.

Mertz, J. A., Simper, M. S., Lozano, M. M., Payne, S. M., & Dudley, J. P. 2005. Mouse mammary tumor virus encodes a self-regulatory RNA export protein and is a complex retrovirus. ***J Virol***, 79(23): 14737-14747.

Mi, S., Lee, X., Li, X., Veldman, G. M., Finnerty, H., Racie, L., LaVallie, E., Tang, X. Y., Edouard, P., Howes, S., Keith, J. C., Jr., & McCoy, J. M. 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. ***Nature***, 403(6771): 785-789.

Miller, A. D. 2008. Hyaluronidase 2 and its intriguing role as a cell-entry receptor for oncogenic sheep retroviruses. ***Semin Cancer Biol***, 18(4): 296-301.

Miller, D. G., Edwards, R. H., & Miller, A. D. 1994. Cloning of the cellular receptor for amphotropic murine retroviruses reveals homology to that for gibbon ape leukemia virus. ***Proc Natl Acad Sci U S A***, 91(1): 78-82.

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(1): 78-84.

Miller, D. G., & Miller, A. D. 1993. Inhibitors of retrovirus infection are secreted by several hamster cell lines and are also present in hamster sera. ***J Virol***, 67(9): 5346-5352.

Mitchell, R. S., Beitzel, B. F., Schroder, A. R., Shinn, P., Chen, H., Berry, C. C., Ecker, J. R., & Bushman, F. D. 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. ***PLoS Biol***, 2(8): E234.

Miyazawa, T., Tomonaga, K., Kawaguchi, Y., & Mikami, T. 1994. The genome of feline immunodeficiency virus. ***Arch Virol***, 134(3-4): 221-234.

Mora, M., Napolitano, C., Ortega, R., Poulin, E., & Pizarro-Lucero, J. 2015. Feline immunodeficiency virus and feline leukemia virus infection in free-ranging guignas (Leopardus guigna) and sympatric domestic cats in human perturbed landscapes on Chiloe Island, Chile. ***J Wildl Dis***, 51(1): 199-208.

Mun, S., Lee, J., Kim, Y. J., Kim, H. S., & Han, K. 2014. Chimpanzee-specific endogenous retrovirus generates genomic variations in the chimpanzee genome. ***PloS One***, 9(7): e101195.

Munson, L., Marker, L., Dubovi, E., Spencer, J. A., Evermann, J. F., & O'Brien, S. J. 2004. Serosurvey of viral infections in free-ranging Namibian cheetahs (Acinonyx jubatus). ***J Wildl Dis***, 40(1): 23-31.

Mura, M., Murcia, P., Caporale, M., Spencer, T. E., Nagashima, K., Rein, A., & Palmarini, M. 2004. Late viral interference induced by transdominant Gag of an endogenous retrovirus. ***Proc Natl Acad Sci U S A***, 101(30): 11117-11122.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., & Honjo, T. 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. ***Cell***, 102(5): 553-563.

Naidenko, S. V., Pavlova, E. V., & Kirilyuk, V. E. 2014. Detection of seasonal weight loss and a serologic survey of potential pathogens in wild Pallas' cats (Felis [Otocolobus] manul) of the Daurian Steppe, Russia. ***J Wildl Dis***, 50(2): 188-194.

Narezkina, A., Taganov, K. D., Litwin, S., Stoyanova, R., Hayashi, J., Seeger, C., Skalka, A. M., & Katz, R. A. 2004. Genome-wide analyzes of avian sarcoma virus integration sites. ***J Virol***, 78(21): 11656-11663.

Nisole, S., Stoye, J. P., & Saib, A. 2005. TRIM family proteins: retroviral restriction and antiviral defence. ***Nat Rev Microbiol***, 3(10): 799-808.

Nolen, R. S. 2004. Feline leukemia virus threatens endangered panthers. ***J Am Vet Med Assoc***, 224(11): 1721-1722.

O'Hara, B., Johann, S. V., Klinger, H. P., Blair, D. G., Rubinson, H., Dunn, K. J., Sass, P., Vitek, S. M., & Robins, T. 1990. Characterization of a human gene conferring sensitivity to infection by gibbon ape leukemia virus. ***Cell Growth Differ***, 1(3): 119-127.

OhAinle, M., Kerns, J. A., Malik, H. S., & Emerman, M. 2006. Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. ***J Virol***, 80(8): 3853-3862.

Oliveira, N. M., Farrell, K. B., & Eiden, M. V. 2006. In vitro characterization of a koala retrovirus. ***J Virol***, 80(6): 3104-3107.

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(3): 371-389, table of contents.

Perez-Caballero, D., Zang, T., Ebrahimi, A., McNatt, M. W., Gregory, D. A., Johnson, M. C., & Bieniasz, P. D. 2009. Tetherin inhibits HIV-1 release by directly tethering virions to cells. ***Cell***, 139(3): 499-511.

Pertel, T., Hausmann, S., Morger, D., Zuger, S., Guerra, J., Lascano, J., Reinhard, C., Santoni, F. A., Uchil, P. D., Chatel, L., Bisiaux, A., Albert, M. L., Strambio-De-Castillia, C., Mothes, W., Pizzato, M., Grutter, M. G., & Luban, J. 2011. TRIM5 is an innate immune sensor for the retrovirus capsid lattice. ***Nature***, 472(7343): 361-365.

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(1-2): 1-7.

Planelles, V., & Barker, E. 2010. Roles of Vpr and Vpx in modulating the virus-host cell relationship. ***Mol Aspects Med***, 31(5): 398-406.

Poleshko, A., Einarson, M. B., Shalginskikh, N., Zhang, R., Adams, P. D., Skalka, A. M., & Katz, R. A. 2010. Identification of a functional network of human epigenetic silencing factors. ***J Biol Chem***, 285(1): 422-433.

Powell, R. D., Holland, P. J., Hollis, T., & Perrino, F. W. 2011. Aicardi-Goutieres syndrome gene and HIV-1 restriction factor SAMHD1 is a dGTP-regulated deoxynucleotide triphosphohydrolase. ***J Biol Chem***, 286(51): 43596-43600.

Priester, W. A., & Mantel, N. 1971. Occurrence of tumors in domestic animals. Data from 12 United States and Canadian colleges of veterinary medicine. ***J Natl Cancer Inst***, 47(6): 1333-1344.

Quackenbush, S. L., Holzschu, D. L., Bowser, P. R., & Casey, J. W. 1997. Transcriptional analysis of walleye dermal sarcoma virus (WDSV). ***Virology***, 237(1): 107-112.

Quigley, J. G., Burns, C. C., Anderson, M. M., Lynch, E. D., Sabo, K. M., Overbaugh, J., & Abkowitz, J. L. 2000. Cloning of the cellular receptor for feline leukemia virus subgroup C (FeLV-C), a retrovirus that induces red cell aplasia. ***Blood***, 95(3): 1093-1099.

Rafatpanah, H., Farid, R., Golanbar, G., & Jabbari Azad, F. 2006. HTLV-I Infection: virus structure, immune response to the virus and genetic association studies in HTLV-I-infected individuals. ***Iran J Allergy Asthma Immunol***, 5(4): 153-166.

Rasheed, S., & Gardner, M. B. 1981. Isolation of feline leukemia virus from a leopard cat cell line and search for retrovirus in wild felidae. ***J Natl Cancer Inst***, 67(4): 929-933.

Rein, A. 2011. Murine leukemia viruses: objects and organisms. ***Adv Virol***, 2011: 403419.

Rice, G. I., Rodero, M. P., & Crow, Y. J. 2015. Human Disease Phenotypes Associated With Mutations in TREX1. ***Journal of Clinical Immunology***, 35(3): 235-243.

Richards, S. 2015. It's more than stamp collecting: how genome sequencing can unify biological research. ***Trends Genet***, 31(7): 411-421.

Robinson, H. L., Astrin, S. M., Senior, A. M., & Salazar, F. H. 1981. Host Susceptibility to endogenous viruses: defective, glycoprotein-expressing proviruses interfere with infections. ***J Virol***, 40(3): 745-751.

Rosenberg, B. R., Hamilton, C. E., Mwangi, M. M., Dewell, S., & Papavasiliou, F. N. 2011. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs. ***Nat Struct Mol Biol***, 18(2): 230-236.

Rous, P. 1910. A Transmissible Avian Neoplasm. (Sarcoma of the Common Fowl.). ***J Exp Med***, 12(5): 696-705.

Rouzine, I. M., Rodrigo, A., & Coffin, J. M. 2001. Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. ***Microbiol Mol Biol Rev***, 65(1): 151-185.

Ruboyianes, R., & Worobey, M. 2016. Foamy-like endogenous retroviruses are extensive and abundant in teleosts. ***Virus Evol***, 2(2): vew032.

Sakaguchi, S., Shojima, T., Fukui, D., & Miyazawa, T. 2015. A soluble envelope protein of endogenous retrovirus (FeLIX) present in serum of domestic cats mediates infection of a pathogenic variant of feline leukemia virus. ***J Gen Virol***, 96(Pt 3): 681-687.

Santoni, F. A., Guerra, J., & Luban, J. 2012. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. ***Retrovirology***, 9: 111.

Sanz-Ramos, M., & Stoye, J. P. 2013. Capsid-binding retrovirus restriction factors: discovery, restriction specificity and implications for the development of novel therapeutics. ***J Gen Virol***, 94(Pt 12): 2587-2598.

Sasada, A., Takaori-Kondo, A., Shirakawa, K., Kobayashi, M., Abudu, A., Hishizawa, M., Imada, K., Tanaka, Y., & Uchiyama, T. 2005. APOBEC3G targets human T-cell leukemia virus type 1. ***Retrovirology***, 2: 32.

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(1): 139-142.

Shapiro, B., Ho, S. Y., Drummond, A. J., Suchard, M. A., Pybus, O. G., & Rambaut, A. 2011. A Bayesian phylogenetic method to estimate unknown sequence ages. ***Mol Biol Evol***, 28(2): 879-887.

Sheets, R. L., Pandey, R., Jen, W. C., & Roy-Burman, P. 1993. Recombinant feline leukemia virus genes detected in naturally occurring feline lymphosarcomas. ***J Virol***, 67(6): 3118-3125.

Schneider, I. C., Eckhardt, M., Brynza, J., Collins, M. K., Cichutek, K., & Buchholz, C. J. 2011. Escape from R-peptide deletion in a gamma-retrovirus. ***Virology***, 418(2): 85-92.

Schroder, A. R., Shinn, P., Chen, H., Berry, C., Ecker, J. R., & Bushman, F. 2002. HIV-1 integration in the human genome favors active genes and local hotspots. ***Cell***, 110(4): 521-529.

Simmons, G., Clarke, D., McKee, J., Young, P., & Meers, J. 2014. Discovery of a novel retrovirus sequence in an Australian native rodent (Melomys burtoni): a putative link between gibbon ape leukemia virus and koala retrovirus. ***PLoS One***, 9(9): e106954.

SJ, O. B., Haussler, D., & Ryder, O. 2014. The birds of Genome10K. ***Gigascience***, 3(1): 32.

Skibba, R. 2016. Geneticists hope to unlock secrets of bats' complex sounds. ***Nature***, 539(7630): 481.

Slater, G. J., Cui, P., Forasiepi, A. M., Lenz, D., Tsangaras, K., Voirin, B., de Moraes-Barros, N., MacPhee, R. D., & Greenwood, A. D. 2016. Evolutionary Relationships among Extinct and Extant Sloths: The Evidence of Mitogenomes and Retroviruses. ***Genome Biol Evol***, 8(3): 607-621.

Sleeman, J. M., Keane, J. M., Johnson, J. S., Brown, R. J., & Woude, S. V. 2001. Feline leukemia virus in a captive bobcat. ***J Wildl Dis***, 37(1): 194-200.

Smart, J. E., Oppermann, H., Czernilofsky, A. P., Purchio, A. F., Erikson, R. L., & Bishop, J. M. 1981. Characterization of sites for tyrosine phosphorylation in the transforming protein of

Rous sarcoma virus (pp60v-src) and its normal cellular homologue (pp60c-src). ***Proc Natl Acad Sci U S A***, 78(10): 6013-6017.

Smith, E. J., Brojatsch, J., Naughton, J., & Young, J. A. 1998. The CAR1 gene encoding a cellular receptor specific for subgroup B and D avian leukosis viruses maps to the chicken tvb locus. ***J Virol***, 72(4): 3501-3503.

Smith, M. F., & Patton, J. L. 1991. Variation in mitochondrial cytochrome b sequence in natural populations of South American akodontine rodents (Muridae: Sigmodontinae). ***Mol Biol Evol***, 8(1): 85-103.

Soll, S. J., Neil, S. J., & Bieniasz, P. D. 2010. Identification of a receptor for an extinct virus. ***Proc Natl Acad Sci U S A***, 107(45): 19496-19501.

Steckbeck, J. D., Kuhlmann, A. S., & Montelaro, R. C. 2014. Structural and functional comparisons of retroviral envelope protein C-terminal domains: still much to learn. ***Viruses***, 6(1): 284-300.

Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J., & Harris, R. S. 2010. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. ***Nat Struct Mol Biol***, 17(2): 222-229.

Stewart, A. F. 2002. Identification of human homologues of the mouse mammary tumor virus receptor. ***Arch Virol***, 147(3): 577-581.

Suzuki, Y., Kobayashi, Y., Horie, M., & Tomonaga, K. 2014. Origin of an endogenous bornavirus-like nucleoprotein element in thirteen-lined ground squirrels. ***Genes Genet Syst***, 89(3): 143-148.

Takeda, A., & Matano, T. 2007. Inhibition of infectious murine leukemia virus production by Fv-4 env gene products exerting dominant negative effect on viral envelope glycoprotein. ***Microbes Infect***, 9(14-15): 1590-1596.

Takeda, E., Nakagawa, S., Nakaya, Y., Tanaka, A., Miyazawa, T., & Yasuda, J. 2012. Identification and functional analysis of three isoforms of bovine BST-2. ***PLoS One***, 7(7): e41483.

Takeuchi, Y., Vile, R. G., Simpson, G., O'Hara, B., Collins, M. K., & Weiss, R. A. 1992. Feline leukemia virus subgroup B uses the same cell surface receptor as gibbon ape leukemia virus. ***J Virol***, 66(2): 1219-1222.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. ***Mol Biol Evol***, 28(10): 2731-2739.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. ***Mol Biol Evol***, 30(12): 2725-2729.

Tarlinton, R., Meers, J., Hanger, J., & Young, P. 2005. Real-time reverse transcriptase PCR for the endogenous koala retrovirus reveals an association between plasma viral load and neoplastic disease in koalas. ***J Gen Virol***, 86(Pt 3): 783-787.

Tarlinton, R. E., Meers, J., & Young, P. R. 2006. Retroviral invasion of the koala genome. ***Nature***, 442(7098): 79-81.

Tonjes, R. R., & Niebert, M. 2003. Relative age of proviral porcine endogenous retrovirus sequences in Sus scrofa based on the molecular clock hypothesis. ***J Virol***, 77(22): 12363-12368.

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): 1-9.

Tsangaras, K., Mayer, J., Alquezar-Planas, D. E., & Greenwood, A. D. 2015. An Evolutionarily Young Polar Bear (Ursus maritimus) Endogenous Retrovirus Identified from Next Generation Sequence Data. ***Viruses***, 7(11): 6089-6107.

Turelli, P., Guiguen, F., Mornex, J. F., Vigne, R., & Querat, G. 1997. dUTPase-minus caprine arthritis-encephalitis virus is attenuated for pathogenesis and accumulates G-to-A substitutions. ***J Virol***, 71(6): 4522-4530.

van de Lagemaat, L. N., Medstrand, P., & Mager, D. L. 2006. Multiple effects govern endogenous retrovirus survival patterns in human gene introns. ***Genome Biol***, 7(9): R86.

van Zeijl, M., Johann, S. V., Closs, E., Cunningham, J., Eddy, R., Shows, T. B., & O'Hara, B. 1994. A human amphotropic retrovirus receptor is a second member of the gibbon ape leukemia virus receptor family. ***Proc Natl Acad Sci U S A***, 91(3): 1168-1172.

Varma, M., Rudolph, D. L., Knuchel, M., Switzer, W. M., Hadlock, K. G., Velligan, M., Chan, L., Foung, S. K., & Lal, R. B. 1995. Enhanced specificity of truncated transmembrane protein for serologic confirmation of human T-cell lymphotropic virus type 1 (HTLV-1) and HTLV-2 infections by western blot (immunoblot) assay containing recombinant envelope glycoproteins. ***J Clin Microbiol***, 33(12): 3239-3244.

Versteeg, G. A., Benke, S., Garcia-Sastre, A., & Rajsbaum, R. 2014. InTRIMsic immunity: Positive and negative regulation of immune signaling by tripartite motif proteins. ***Cytokine Growth Factor Rev***, 25(5): 563-576.

Weiss, R. A. 2006. The discovery of endogenous retroviruses. ***Retrovirology***, 3: 67.

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(10): e13432.

Wu, T., Yan, Y., & Kozak, C. A. 2005. Rmcf2, a xenotropic provirus in the Asian mouse species Mus castaneus, blocks infection by polytropic mouse gammaretroviruses. ***J Virol***, 79(15): 9677-9684.

Wu, X., Li, Y., Crise, B., & Burgess, S. M. 2003. Transcription start regions in the human genome are favored targets for MLV integration. ***Science***, 300(5626): 1749-1751.

Wu, Y., Koharudin, L. M., Mehrens, J., DeLucia, M., Byeon, C. H., Byeon, I. J., Calero, G., Ahn, J., & Gronenborn, A. M. 2015. Structural Basis of Clade-specific Engagement of SAMHD1 (Sterile alpha Motif and Histidine/Aspartate-containing Protein 1) Restriction Factors by Lentiviral Viral Protein X (Vpx) Virulence Factors. ***J Biol Chem***, 290(29): 17935-17945.

Xu, W., Gorman, K., Santiago, J. C., Kluska, K., & Eiden, M. V. 2015. Genetic diversity of koala retroviral envelopes. ***Viruses***, 7(3): 1258-1270.

Yan, Y., Buckler-White, A., Wollenberg, K., & Kozak, C. A. 2009. Origin, antiviral function and evidence for positive selection of the gammaretrovirus restriction gene Fv1 in the genus Mus. ***Proc Natl Acad Sci U S A***, 106(9): 3259-3263.

Yang, L., Guell, M., Niu, D., George, H., Lesha, E., Grishin, D., Aach, J., Shrock, E., Xu, W., Poci, J., Cortazio, R., Wilkinson, R. A., Fishman, J. A., & Church, G. 2015. Genome-wide inactivation of porcine endogenous retroviruses (PERVs). ***Science***, 350(6264): 1101-1104.

Yap, M. W., Colbeck, E., Ellis, S. A., & Stoye, J. P. 2014. Evolution of the retroviral restriction gene Fv1: inhibition of non-MLV retroviruses. ***PLoS Pathog***, 10(3): e1003968.

Yap, M. W., & Stoye, J. P. 2013. Apparent effect of rabbit endogenous lentivirus type K acquisition on retrovirus restriction by lagomorph Trim5alphas. ***Philos Trans R Soc Lond B Biol Sci***, 368(1626): 20120498.

Ye, L., Bu, Z., Vzorov, A., Taylor, D., Compans, R. W., & Yang, C. 2004. Surface stability and immunogenicity of the human immunodeficiency virus envelope glycoprotein: role of the cytoplasmic domain. ***J Virol***, 78(24): 13409-13419.

Zeng, M., Hu, Z., Shi, X., Li, X., Zhan, X., Li, X. D., Wang, J., Choi, J. H., Wang, K. W., Purrington, T., Tang, M., Fina, M., DeBerardinis, R. J., Moresco, E. M., Pedersen, G., McInerney, G. M., Karlsson Hedestam, G. B., Chen, Z. J., & Beutler, B. 2014. MAVS, cGAS, and endogenous retroviruses in T-independent B cell responses. ***Science***, 346(6216): 1486-1492.

Zhang, G., Rahbek, C., Graves, G. R., Lei, F., Jarvis, E. D., & Gilbert, M. T. 2015. Genomics: Bird sequencing project takes off. ***Nature***, 522(7554): 34.

Zou, L., Peng, Q., Wang, P., & Zhou, B. 2016. Progress in Research and Application of HIV-1 TAT-Derived Cell-Penetrating Peptide. ***J Membr Biol***.

9 Supplement

Publication 1: Endogenous Lentivirus in Malayan colugo (Galeopterus variegatus), a close relative of primates

Publication 2: Life History of the Oldest Lentivirus: Characterization of ELVgv Integrations in the Dermopteran Genome

Publication 3: Discovery of the first endogenous Deltaretrovirus, in the genome of long-fingered bats (Miniopteridae)

Publication 4: Induction and characterization of a replication competent cervid endogenous Gammaretrovirus (CrERV) from mule deer cells