Expression and Characterization of Ancient Retrovirus Envelope Genes

Author: Kate Halm

EXPRESSION AND CHARACTERIZATION OF ANCIENT RETROVIRUS ENVELOPE GENES

Kate Halm

A dissertation

submitted to the Faculty of

the department of Biology

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College
Morrissey College of Arts and Sciences
Graduate School

August 2018

Expression and Characterization of Ancient Retrovirus Envelope Genes

Kate Halm

Advisor: Welkin E. Johnson Ph.D.

Endogenous retroviruses (ERVs) make up a significant portion of vertebrate genomes, and serve as a fossil record of past retroviral infections. Although most ERV genes acquire inactivating mutations over time, some loci retain open reading frames (ORFs) across one or more of the viral genes. The ERV-Fc family, for example, endogenized in multiple mammalian hosts 10 to 30 million years ago, yet many copies maintain intact ORFs corresponding to the *env* gene, including loci in humans (HERV-Fc1-env) and baboons (babERV-Fc2-env). We previously identified intact ERV-Fc-related *env* sequences in eight additional mammalian species: chimpanzee, bonobo, aardvark, grey mouse lemur, squirrel monkey, marmoset, dog, and panda. Here we present the results of our assays of expression of these full-length Env proteins. We found that most of the precursors were not cleaved to form the functional surface (SU) and transmembrane (TM) subunits, even when a canonical furin cleavage site was still intact. An exception was babERV-Fc2, in which reconstruction of the cleavage site led to cleavage into SU and TM subunits. Furthermore, removal of 22 residues from the C-terminus of the cytoplasmic tail of babERV-Fc2 enhanced syncytia formation and the ability of babERV-Fc2 pseudotyped virions to infect 293T cells, suggesting the presence of an R-peptide cleavage mechanism. A survey of a small panel of cells revealed that only human cell lines were infectable by babERV-Fc2 pseudotyped murine leukemia virus (MLV) particles,

whereas cells of old world monkey, canine, feline and chicken origin were not susceptible to infection. Ectopic expression of native Env codon optimized babERV-Fc2 Env can also inhibit infection by reconstructed babERV-Fc2 pseudotyped virus, raising the possibility that the endogenous glycoprotein encoded in the baboon genome may function as a viral entry inhibitor. Our results suggest that exaptation of ERV Env proteins as antiviral defense genes involves a combination of selective pressures: selection to preserve the receptor-binding and receptor interference functions of Env, but also selection to eliminate the membrane fusion related functions.

Acknowledgements

First, I would like to thank my advisor Welkin Johnson for introducing me to the wonderful world of viruses and in particular to endogenous retroviruses. In an odd way, working with them satisfies both my love of science and history. I would also like to thank Welkin for sticking with this project even when it seemed like it was going nowhere, and for having endless advice and ideas for the next experiment to try. I would also like to thank the rest of the Johnson lab, past and present members, for being supportive and helpful in and out of the lab. In particular, I would like to thank Ted Diehl for starting this project and helping me get it off the ground, also Rana Akleh and Anindita Sinha for making the lab a more enjoyable place to work and always being there to talk to. I would like to thank the undergrads that helped me on this project, Ellen White and Josh Elbaz. Additionally, I would like to thank my committee, Ken Williams, Charlie Hoffman, Julia Halo and Ben Fofana for taking the time to give me feedback on my thesis work.

I would like to thank all the people that have helped make my time in Boston and in grad school more enjoyable. This includes the bio grad students that I have met and made friends with, who know how to have a good time outside of the lab and can commiserate and encourage each other when things get tough. In particular, I would like to thank David Layman for being a constant friend and putting up with my cat whenever I was out of town. I would like to thank all of the people who I met through playing inner tube water polo. Playing ITWP was a great way to alleviate some stress, it also allowed me to meet an

amazing group of people that have always been very supportive and interested in my journey through grad school.

Finally, I want to thank my boyfriend Nathan Serrano and parents Cathy Simon and Doug Halm. Nathan for being incredibly supportive during this whole writing process and defense process. We met at a very stressful time in my life and he has been amazing throughout all of it, always encouraging me to work hard but also making sure I take the breaks I need. My Mom, Cathy, I would like to thank for always being my first editor on everything I write. My Dad, Doug, I would like to thank for trying to understand my work and sending me science articles that I might find interesting. Both of my parents have always been so understanding and supportive in everything I have done in my life, and have encouraged my brothers and me to think for ourselves and strive for more knowledge. Because of that I have always loved learning new things and that thirst for knowledge lead me to grad school.

Table of Contents

Page

Page

List of Figures and Tables

Figures

Page

Tables

List of Abbreviations

AIDS	Acquired immune deficiency syndrome
ALV	Avian leucosis virus
ameERV	Panda endogenous retrovirus
ancHTenv	Ancestral HERV-T *env*
babERV	Baboon endogenous retrovirus
babSP	babERV-Fc2 signal peptide
BLASTn	Basic local alignment search tool n
BLV	Bovine leukemia virus
CA	Capsid protein
cfERV	Dog endogenous retrovirus
cjaERV	Marmoset endogenous retrovirus
cl	Indicates a furin cleavage site that has been repaired
conERV	Consensus endogenous retrovirus
DNA	Deoxyribonucleic acid
EIAV	Equine infectious anemia virus
env	Envelope gene
Env	Envelope protein
ER	Endoplasmic reticulum
ERV	Endogenous retrovirus
ERV-DC	Endogenous gammaretrovirus of domestic cats
F-MLV	Friend murine leukemia virus
Fc1SP	conERV-Fc1 signal peptide

FeLV	Feline leukemia virus
FFV	Feline foamy virus
FP	Fusion peptide
Fv-1	Friend virus susceptibility factor 1
Fv-4	Friend virus susceptibility factor 4
gag	Group-specific antigen gene
Gag	Group-specific antigen protein
GaLV	Gibbon ape leukemia virus
gmlERV	Grey mouse lemur endogenous retrovirus
HERV	Human endogenous retrovirus
HIV-1	Human immunodeficiency virus type 1
hMCT1	Human monocarboxylate transporter 1
homERV	Chimpanzee endogenous retrovirus
hsaHTenv	Modern HERV-T *env*
HTLVs	Human T-lymphotropic viruses
IN	Integrase
ISD	Immunosuppressive domain
JSRV	Jaagsiekte sheep retrovirus
KIDS	Koala immune deficiency syndrome
KoRV	Koala retrovirus
LTR	Long terminal repeat
MA	Matrix protein
MFC	Mink cell focus-forming MLV

MLV	Murine leukemia virus
MLVct	Murine leukemia virus cytoplasmic tail
MMTV	Mouse mammary tumor virus
MPMV	Mason-Pfizer monkey virus
mRNA	Messenger RNA
MSD	Membrane spanning domain
MuERV	Murine endogenous retrovirus
NC	Nucleocapsid protein
oafERV	Aardvark endogenous retrovirus
ORF	Open reading frame
PBS	Primer binding sequence
PERV	Porcine endogenous retrovirus
PIC	Pre-integration complex
pol	Polymerase gene
Pol	Polymerase protein
ppaERV	Bonobo endogenous retrovirus
PPT	Polypurine tract
PR	Protease
pro	Protease gene
Pro	Protease protein
R	Repeat sequence
RDR	RD-114 and D-type retrovirus
Refrex-1	Restriction for feline retrovirus X

RNA	Ribonucleic acid
RNase H	Ribonuclease H
RT	Reverse transcriptase
sboERV	Squirrel monkey endogenous retrovirus
SNV	Spleen necrosis virus
SP	Signal peptide
SRV	Simian retrovirus
SU	Surface subunit
TM	Transmembrane subunit
tRNA	Transfer RNA
U3	Unique 3’ sequence
U5	Unique 5’ sequence
VSVG	Vesicular stomatitis virus glycoprotein
WDSV	Walleye dermal sarcoma virus
WEHV	Walleye epidermal hyperplasia virus

Chapter 1: Introduction

Retroviridae

The *Retroviridae* is a family of viruses, known as retroviruses, that are characterized by the ability to reverse transcribe their positive-strand (+) ribonucleic acid (RNA) genome into a deoxyribonucleic acid (DNA) copy, which is then integrated into the host genome. This integrated DNA genome is then known as a provirus. Retroviruses are widespread and can infect many vertebrates. Retroviruses have enveloped virions, with envelope proteins (Env) studding the surface; within there is a capsid core that protects the two copies of the RNA genome and contains the viral enzymes (Fig. 1.1) [1]. All retrovirus RNA genomes have four standard genes: group-specific antigen (*gag*), protease (*pro*), polymerase (*pol*) and envelope (*env*). These genes are flanked on the 5' side by the repeat sequence (R), the unique 5' sequence (U5), the primer binding sequence (PBS) and a packaging signal. On the 3' side they are bordered by a polypurine tract (PPT), the unique 3' sequence (U3) and R (Fig. 1.2 and 1.3) [1]. The RNA genome has a 5' cap and a 3' poly-A tail.

Life Cycle

The retroviral life cycle begins with the Env protein recognizing and binding to a cell surface receptor. The Env is expressed as a polyprotein that is cleaved into two subunits, a surface subunit (SU) and a transmembrane subunit (TM). SU is responsible for receptor recognition. A conformational change is induced either by receptor binding, or for viruses that enter via endocytosis, by a drop in pH, triggering the fusion peptide (FP) motif in the TM subunit which

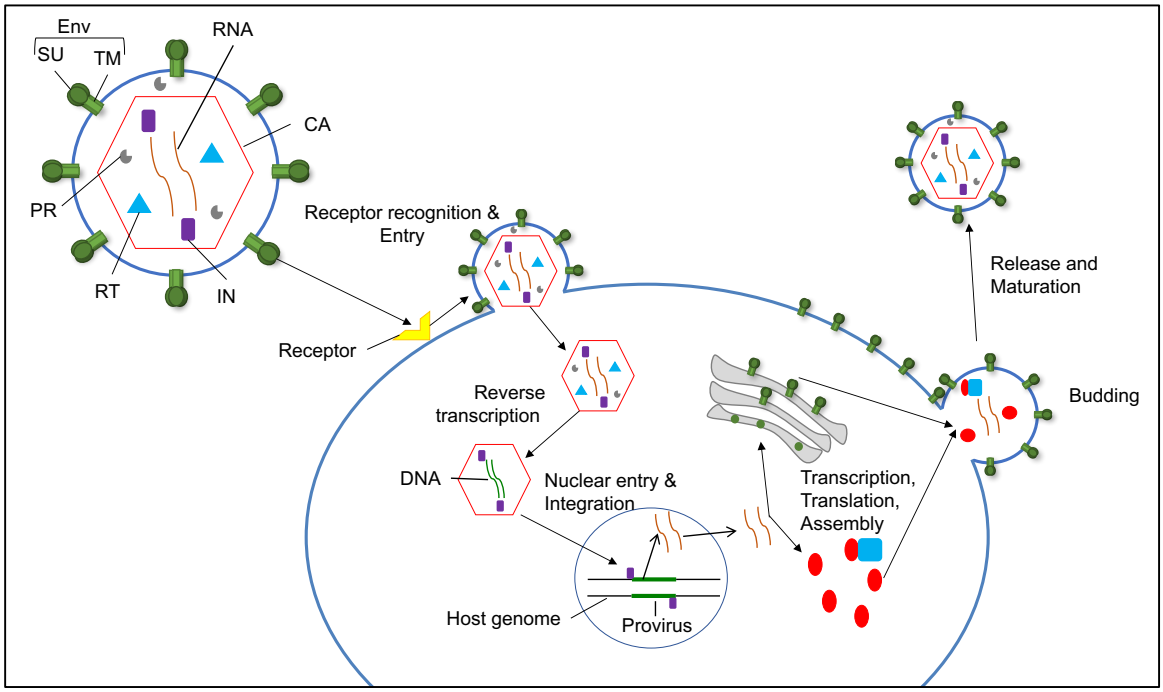


Figure 1.1: The retroviral life cycle. In the top left corner is a representation of a mature retrovirus, CA=capsid (red), IN=integrase (purple), RT=reverse transcriptase (light blue), PR=protease (dark gray), Env=envelope (green) (SU=surface subunit, TM=transmembrane domain). Arrows follow the viral life cycle starting with Env recognition of a receptor and ending with release and maturation of a new viral particle.

becomes inserted into the host cell membrane; fusion then permits entry of the viral core into the cytoplasm (Fig. 1.1) [1], [2]. This fusion process can take place at the cell surface or after endocytosis of the virus. After entry of the viral core into the cytoplasm, reverse transcription of the RNA genome into the DNA genome begins, and takes place in a large complex called the pre-integration complex (PIC). Reverse transcriptase (RT) is the major enzyme involved in this process [1], [3]–[5] (Fig. 1.1). There is a small amount of uncoating of the core that has to occur for reverse transcription to happen, but the process and how much is poorly understood. Reverse transcription starts with binding a primer transfer RNA (tRNA) to a complementary PBS on the plus strand RNA genome (Fig. 1.2). The RT then binds to the primer and minus strand (-) DNA synthesis occurs towards the 5' end producing the U5 and R sequences, the RNA template is degraded by ribonuclease H (RNase H) activity after it is copied into DNA (Fig. 1.2). This fragment of DNA then "jumps" to the 3' end of the viral genome, where the R sequence anneals to the complementary R sequence at the 3' end. DNA synthesis will continue and stop at the PBS (Fig. 1.2). To synthesize the (+) strand of DNA the PPT, which is mostly resistant to the RNase H activity, serves as the primer binding spot (Fig. 1.2). DNA synthesis occurs toward the 5' end, copying the U3, R and U5, the tRNA primer is removed and there is a second translocation step (Fig. 1.2). The DNA then anneals to the complementary PBS on the other side of the strand and RT catalyzes the elongation of the remaining DNA genome. The DNA copy of the RNA genome has identical long terminal

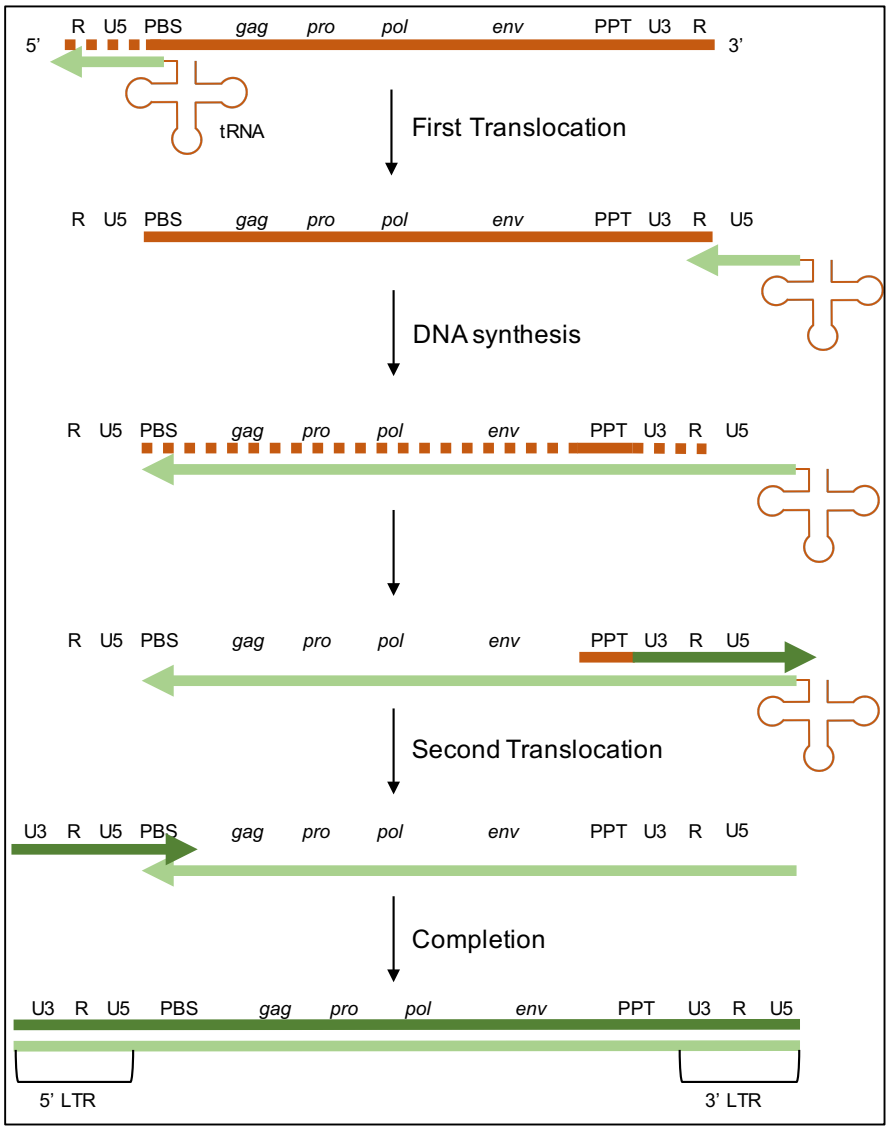


Figure 1.2: Major steps of reverse transcription. Orange lines represent RNA, dashed orange lines indicate degrading RNA and green lines represent DNA. R = repeat region, U5 = unique 5' sequence, PBS = primer binding sequence, *gag* = group-specific antigen, *pro* = protease, *pol* = polymerase, *env* =envelope, PPT = polypurine tract, U3 = unique 3' sequence, LTR = long terminal repeat.

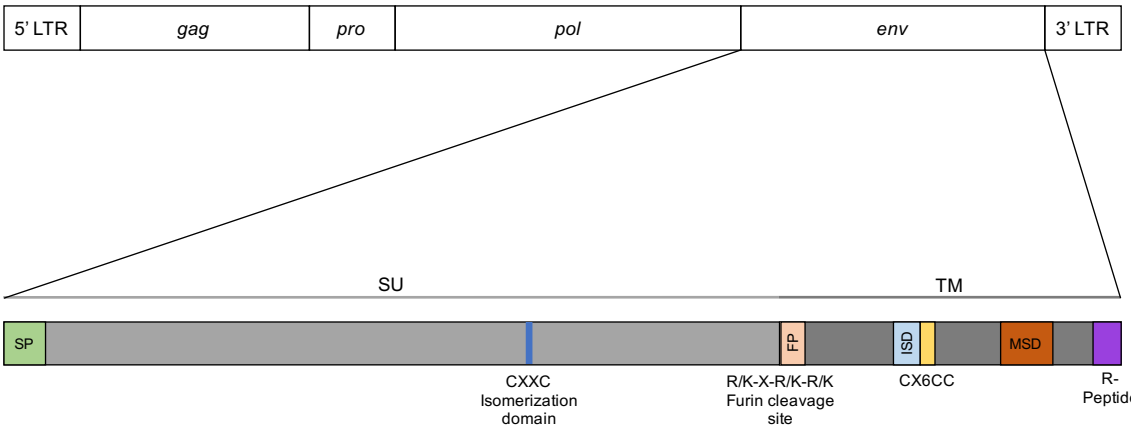


Figure 1.3: Representation of a simple retrovirus provirus and a gamma Env. Top figure is a representation of a simple retroviral provirus with LTR's flanking the *gag, pro, pol* and *env* genes. Blown up below is a typical gammaretrovirus Env. Light gray=surface unit (SU), gray=transmembrane domain (TM), green=signal peptide (SP), blue bar=isomerization domain, light orange=fusion peptide (FP), light blue=immunosuppressive domain (ISD), yellow=CX6CC motif, dark orange=membrane spanning domain (MSD) and purple=R-peptide. The SU and TM are cleaved at the furin cleavage site by furin.

repeats (LTRs) at either end that respectively contain the U3, R and U5 (Fig. 1.2 and 1.3) [1].

After reverse transcription, the DNA genome is trafficked to the nucleus (Fig. 1.1). The mechanism of nuclear entry is not well known, but most retroviruses require the cell to go through mitosis and for the nuclear envelope to breakdown before entry occurs [1], [6]–[9]. Though, lentiviruses can infect non-dividing cells, suggesting they are able to cross an intact nuclear envelope [1], [6], [10], [11]. The DNA viral genome is then integrated into the host genome; this process is mediated by the viral protein integrase (IN) (Fig. 1.1). Integrase removes two nucleotides from the 3' ends of the viral DNA genome creating an overhang. A double strand DNA break, with an overhang, is then introduced in the host genome by IN. A new phosphodiester bond is created between the host genome and the viral genome, the overhanging 5' viral DNA is not integrated into the host DNA. After integration, there are short gaps in the host genome that are repaired by the host DNA repair machinery, resulting in short duplication sites that flank the provirus, which can be anywhere from 4-6bp long depending on the retrovirus [1]. The provirus is then a permanent part of the host cell genome (Fig. 1.1).

Production of new virions from the provirus through transcription and translation is mediated by host enzymes. The U3 contains a promoter that is recognized by RNA polymerase II. Transcription starts at U3-R of the 5' LTR, the RNA is capped and polyadenylated, which generates a new RNA viral genome or messenger RNA (mRNA) to be used as a template for translation the viral

proteins [1]. The RNA genomes are trafficked out of the nucleus and will take several pathways (Fig. 1.1). Some will be trafficked directly to sites of virion assembly, to be used as a new genome. Other copies will be used to produce the Gag and Gag-Pro-Pol polyproteins, and another copy will be spliced to yield mRNA splice variants for the Env protein and accessory proteins, if any are present (Fig. 1.1). The viral proteins produced will be trafficked to the cell membrane where viral assembly typically occurs (Fig. 1.1). The Gag, Pro and Pol proteins are expressed in a complex manner which differs between the various types of retroviruses; they are translated into a polyprotein that is processed during the maturation stage of the viral life cycle. Because more Gag is needed than Pro or Pol, retroviruses have evolved two main mechanisms to limit the amount of the polyprotein produced. The first is translational readthrough, in which the *gag, pro* and *pol* ORFs are in the same reading frame but are separated by a stop codon. In most instances translation only produces Gag protein. About 5% to 10% of the time translation does not cease at the stop codon but instead continues to translate the RNA through the entire length to produce the Gag-Pro-Pol precursor protein [1]. The second mechanism used by retroviruses to regulate the ratios of Gag, Pro and Pol is through translational frameshifting. This occurs when the *gag, pro* and *pol* genes are in different reading frames. Most occurrences result in only Gag being produced, but approximately 10% of the time the ribosome will slip back a nucleotide and continue in the new reading frame to produce the Gag-Pro-Pol polyprotein [1]. The *pro* reading frame, depending on the virus, can be part of *gag*, part of *pol* or

in its own reading frame. The production of Gag, Pro and Pol occurs in the cytoplasm and the proteins are then trafficked to the cell membrane (Fig. 1.1).

The Env protein is produced from a spliced variant of the viral mRNA. Env contains a hydrophobic signal peptide (SP) within the first ~20 amino acids that directs translation of the *env* to the endoplasmic reticulum (ER) and through the secretory pathway (Fig.1.1). The SP is cleaved off by cellular proteins and in the ER the Env is folded and oligomerized; most Envs must be formed into trimers to function properly [1]. The Env protein is subsequently exported to the Golgi where it is cleaved into its two subunits SU and TM by the cellular enzyme furin, which recognizes the motif R/K-X-K/R-R/K. The two Env subunits can be covalently or noncovalently associated depending on the type of virus. In the Golgi, Env proteins are usually heavily glycosylated, most commonly through N-linked glycosylation. Env is then trafficked to the cell surface, where it is anchored into the cell membrane by the membrane spanning domain (MSD) located in the TM (Fig. 1.1) [1].

Once viral proteins have been expressed and transported to the cell membrane, progeny virions will be produced. The Gag precursor is largely responsible for driving particle assembly. Gag proteins assemble at the membrane and begin to curve and bud out; they eventually form a sphere that is connected to the cell by a narrow stalk that is pinched off [1]. This immature virus is spherical and contains the RNA genome and all precursor proteins (Fig. 1.1). Alternatively, some retroviruses assemble in the cytoplasm and are transported to the membrane to bud out of the cell [1]. As the retrovirus is

budding from the cell, the precursor Gag and Gag-Pro-Pol proteins are cleaved by the viral protease (PR). Gag is cleaved into the matrix protein (MA) which stays associated with the membrane surrounding the virion, and the capsid protein (CA), which forms the viral core (which may be spherical, cylindrical or conical depending on the virus type). The core contains the RNA genome and the proteins required for reverse transcription and integration. Gag cleavage also produces the nucleocapsid protein (NC), which is closely associated with the two copies of the RNA genome and coats both copies. NC has additionally been shown to be involved with reverse transcription and integration [1]. While Gag is cleaved, the Gag-Pro-Pol polyprotein is cleaved into its various products including the viral protease (PR), the reverse transcriptase (RT) and integrase (IN). Depending on the type of virus these can form a diverse pattern. After release from the cell and cleavage of the precursor proteins, there is a morphological change in the virions where they typically become more compact (Fig. 1.1) [1]. Once this morphological change occurs, the mature virion can infect a new host cell.

Classification of Retroviruses

The *Retroviridae* are currently grouped into seven distinct genera: Alpharetroviruses, Betaretroviruses, Gammaretroviruses, Deltaretroviruses, Epsilonretroviruses, Lentiviruses and Spumaviruses, these classifications are generally based on their reverse transcriptase gene, core structure and genome [1].

Alpharetroviruses

Alpharetroviruses are simple retroviruses that do not encode accessory proteins. They have a C-type morphology, meaning they assemble at the plasma membrane and have a spherical centrally placed core. During translation of the Gag-Pro-Pol polyprotein translational frameshifting is used. *Pro* is located at the 3' end of *gag* and is in the same frame, while *pol* is in a different reading frame. The TM and SU subunits of the Env protein remain covalently associated after cleavage. Alpharetroviruses infect birds; a prototypical alpharetrovirus is avian leukosis virus (ALV) that infects chickens and can cause various cancers in the avian host [12].

Betaretroviruses

Betaretroviruses are simple retroviruses that infect mammals including mice, sheep and primates. They have either a B-type morphology with round non-centrally placed core or a D-type morphology with a cylindrical core. *Gag*, *pro*, and *pol* are in different reading frames; translational frameshifting is used. Immature betaretroviral particle assemble in the cytoplasm and are transported to the cell membrane to bud out of the cell [1]. The two subunits of Env, SU and TM, are non-covalently associated [1], [13]. Examples of betaretroviruses include mouse mammary tumor virus (MMTV) which causes cancerous growth in mammary glands of mice and can be passed to offspring through the milk or in an endogenous fashion [14]. MMTV has an additional assessor gene called *sag,* or superantigen, that causes T-cells to be activated [1]. Jaagsiekte sheep retrovirus (JSRV) infects sheep and causes lung cancer

[15]. Mason-Pfizer monkey virus (MPMV), which is part of a larger group of viruses called simian retroviruses (SRV), infects Asian macaques causing immune deficiency that can be fatal [16]. Interestingly the MPMV *env* originated from a crossover event and is of gammaretroviral origin [13].

Gammaretroviruses

Gammaretroviruses are simple retroviruses with the largest number of known members, potentially due to the wide species tropism observed, infecting mammals, reptiles and birds. They have a C-type morphology assembling at the cell surface with a spherical central core. Gag, Pro and Pol are produced through translational readthrough with a stop codon at the end of *gag*. Gammaretroviral Env subunits are covalently associated and commonly recognize a multimembrane spanning protein as a receptor [1], [13], [17]. Gamma Envs commonly require processing in the cytoplasmic tail by the viral protease during the maturation stage in order to activate the fusogenic ability of Env (Fig. 1.3) [18]–[23].There are numerous examples of gammaretrovirus, three of them are: murine leukemia virus (MLV), feline leukemia virus (FeLV) and koala retrovirus (KoRV). MLV can cause cancer in mice, FeLV has many detrimental health effects in cats but the final stage is the development of lymphomas, KoRV has been implicated in koala immune deficiency syndrome (KIDS) which can lead to many health risks for koalas [24]–[26]. All three of these viruses have exogenous and endogenous forms in their respective hosts.

Deltaretroviruses

Deltaretroviruses are complex retroviruses that have several additional genes including *rex* and *tax* which are involved with the synthesis and processing of the viral RNAs. These retroviruses have a C-type morphology assembling at the cell surface and have a spherical central core. They use translational frameshifting; the *gag*, *pro*, and *pol* are all in different reading frames [1]. They have gamma-type Envs, with SU and TM covalently associated [13]. Deltaretroviruses are known to infect primates and bovine, examples being human T-lymphotropic viruses (HTLVs) and bovine leukemia virus (BLV), they can cause leukemia in their host along with other health problems [27], [28].

Epsilonretroviruses

Epsilonretroviruses are complex retroviruses with one to three additional open reading frames termed *ORFs a, b* or *c*. They have a C-type morphology. Translational readthrough is used in translation with a stop codon at the end of *gag* [1]. their Envs are not well characterized but sequence analysis suggests a unique morphology [29]. Epsilonretroviruses are waterborne and infect mainly fish. Two examples are walleye dermal sarcoma virus (WDSV), which causes tumors in walleye; and walleye epidermal hyperplasia virus 1 and 2 (WEHV-1 or 2) which induce a neoplastic condition in fish resulting in lesions [30], [31].

Lentiviruses

Lentiviruses are complex retroviruses that infect a wide variety of mammals including primates, cats, horses, cows and rabbits [1]. They can have multiple accessory genes specific to the virus, which have numerous functions

such as: antagonizing the host innate immune system, controling transcription, RNA processing, virion assembly and host gene expression [1]. During infection lentiviruses are able to infect dividing and non-dividing cells [6], [10], [11]. The capsid core has a unique morphology being either cylindrical or conical. During translation, translational frameshifting is used. The *gag* is in its own reading frame and the *pro-pol* is in a different reading frame. The Env subunits are non-covalently associated in lentiviruses. There are many examples of lentiviruses. Two examples are: human immunodeficiency virus type 1 (HIV-1), which if left untreated it can lead to acquired immune deficiency syndrome (AIDS) [1] and equine infectious anemia virus (EIAV), which can cause multiple health complications in horses, the main one being anemia [32].

Spumaviruses

Spumaviruses (also known as foamy viruses) are complex retroviruses, unique in multiple different ways, including their morphology characterized by large surface spikes and an uncondensed core [1], [33]. Virion assembly occurs in the cytoplasm and budding takes place either from the ER or the plasma membrane. Translation of *gag* and *pol* is unique because *pol* is translated from a mRNA splice variant. At the 3' end of *env* there is another transcriptional start site [1], [33]. Spumaviruses characteristically form large vacuoles in their host. Infection by a foamy virus is usually nonpathogenic [33].

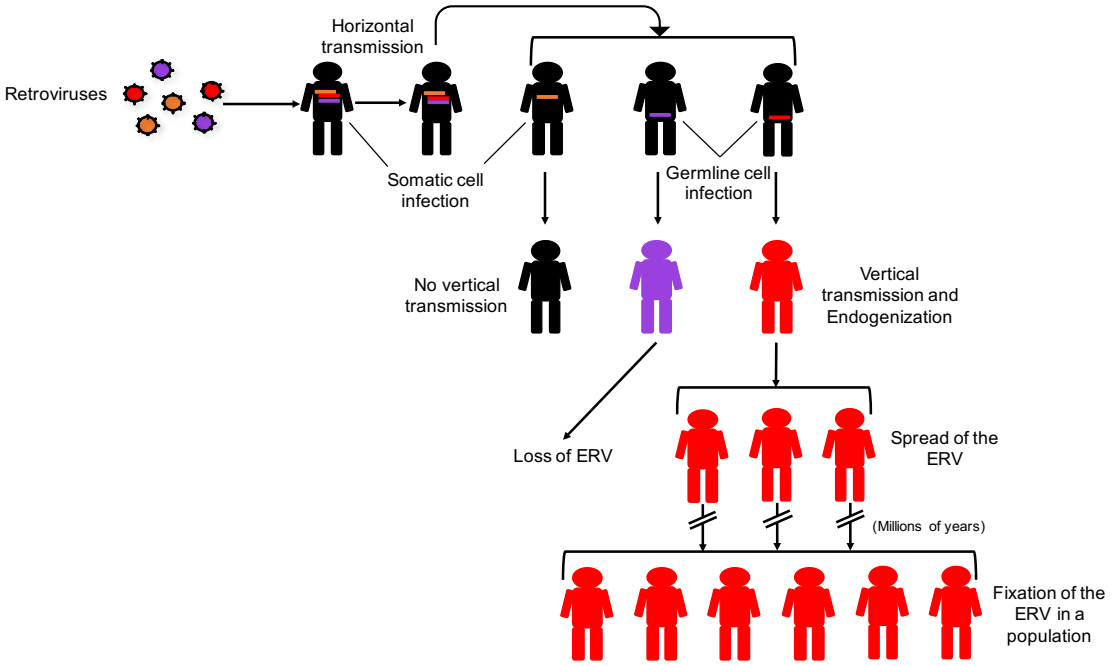


Figure 1.4: Endogenization of a Retrovirus. Spreading retroviruses infect a host by inserting their genome into the host DNA, most commonly into a somatic cell. Production of virus from these somatic cells results in horizontal transmission of the virus within a population (top left). Occasionally, a retrovirus will infect a germline cell (purple and red virus), when this occurs the proviral DNA can be vertically transmitted to the offspring, somatic cell infections cannot be vertically transmitted to the offspring (orange virus). When vertical transmission happens, the retrovirus has been endogenized and is present in all the somatic and germline cells of the host. Several events can happen once a retrovirus is endogenized, perhaps the most common event is for the ERV to be lost (purple virus). If not lost an ERV will be passed on in a Mendelian fashion spreading in the population (red virus middle) and eventually over millions of years it can become fixed within a population (red virus bottom).

Endogenous Retroviruses

The normal mode of spread of an exogenous virus is horizontally from one host to another, usually infecting somatic cells (Fig. 1.4). Occasionally a retrovirus will infect a host germline cell and the provirus can then be passed vertically to the host's offspring (Fig. 1.4). The retroviral provirus is then present in every cell of the offspring and is considered endogenized creating a new endogenous retrovirus (ERV) (Fig. 1.4). ERVs are passed on in a Mendelian fashion, though, most new ERVs are lost soon after insertion, which could be due to it being detrimental to the host or the host lineage being lost (Fig. 1.4). Occasionally an ERV will become fixed within a population, which can take millions of years (Fig. 1.4). Once part of the host genome, the ERV provirus is subject to the host mutation rate and typically evolves with the host as any other gene. In most cases, the ERV provirus accumulates deactivating mutations such that infectious virions can no longer be produced. Another common occurrence is the formation of solo LTRs. Because LTRs are identical upon integration, recombination between the two can occur, which results in loss of the ERV proviral genome between the LTRs, leaving one LTR in its place [34]–[36]. There are a couple of ways to calculate the age of ERVs. First, because the LTRs are identical upon insertion, the number of accumulated mutations between the two can be used to estimate the age of the ERV. By applying the host mutation rate and the divergence between the two LTRs an estimated time of insertion can be calculated [34], [37]–[41]. Additionally, if an ERV is present at the same locus in different species (orthologous insertions), the insertion event

must have predated speciation, and by utilizing known speciation timelines the ERVs age can be estimated [34],[42].

Typically, only one copy of a retrovirus enters the germline, but once there the copies of the ERV may be expressed and reinserted into the host genome, which can lead to an expansion of the ERV within the genome resulting in multiple copies [34], [42]. Deactivating mutations seem to be the major form of controlling expression of ERVs, however, there is some question of how potentially active ERVs are controlled. The main host method of controlling ERV expression appears to be epigenetic, through methylation modification in order to deactivate the ERV [34], [42]–[45]. Occasionally, an ERV proviral gene will have a maintained open reading frame (ORF). It is currently not known why most of these genes have ORFs, although in some instances these ERV ORFs have been co-opted by the host for a function (see below).

ERVs were discovered and characterized in the late 1960s and early 1970s. One of the first studied was endogenous ALV; ALV was a problem in chickens in the 1960s, to find a solution to the virus a serological test was developed to test for ALV Gag. However, it was discovered that some uninfected chickens still tested positive for ALV [46], [47]. It was further shown that the Gag antigen that was being detected was inherited in a Mendelian fashion [48], [49]. Around the same time, it was observed that some chicken embryo cells would release infectious pseudotyped virus without an Env being provided [46], [50], [51]. This suggested the possibility of an endogenous *env* [46], [52], [53]. It was observed with inbred chickens, that the *env* was being inherited in a

Mendelian fashion [54]. With the discovery of RT, uninfected chickens were tested for proviral ALV DNA. The test revealed numerous copies of ALV in most chicken breeds, many of them defective. Further tests confirmed the finding and identified differences between the endogenous and exogenous retroviruses [8], [42], [46], [55]–[59]. While endogenous ALV was being discovered, endogenous MLV was also detected. In 1933, a mouse strain (AKR) was developed that had a high probability of developing lymphoma, however the cause was unknown at the time [7],[60], [61]. When MLV was classified as a virus in 1951, it was also understood to be the causative agent in the formation of lymphomas in AKR mice [46], [60]. It was not until the 1970's that it was discovered it was endogenous MLV causing lymphomas in AKR mice. The same as endogenous ALV, uninfected mouse cells were observed to spontaneously release MLV. It was also found that there were numerous copies of endogenous MLV that were largely defective [42], [46], [62]–[64]. The discovery of these ERVs opened a new line of investigation, and through the use of multiple different methods including hybridization and PCR, other ERVs were discovered [65], [66]. However, it was not until full genome sequencing that the full extent of ERVs could be appreciated. ERVs are widespread throughout vertebrate animals; for example, in humans they comprise about 8% of the genome, and other sequenced vertebrate genomes show similar numbers across the board [67]–[69].

While it is easy to think of ERVs as ancient, there are examples of ERVs that are polymorphic within populations and that are still in the process of endogenizing, meaning they are endogenized but are not fixed. An example is

human ERV-K (HERV-K), which began infecting germlines and endogenizing before the split from old world monkeys 25-30 million years ago [2]. However, HERV-K appears to have been active fairly recently as evidenced by an insertion that seems to be intact and by the fact it is polymorphic within the human population [2], [70]–[72]. Another example of an ERV that is still in the process of endogenizing is KoRV-A, which is currently spreading both horizontally and vertically in koalas [73].

ERV-Fc Family of ERVs

The work described in this thesis involves *env* genes belonging to ERVs of the ERV-Fc family. ERV-Fc is a gamma-like endogenous retrovirus that was first characterized in the Heidmann lab [74], [75]. During a screen of genomic databases using the immunosuppressive domain (ISD) as a query, an ERV Env was found in the human genome that did not fit into any known ERV families. Due to its phenylalanine (F) tRNA PBS it was named human ERV-Fc (HERV-Fc) [75]. HERV-Fc is scientifically important as it has a maintained *env* ORF, consequently the Heidmann lab looked further into HERV-Fc. They found that in humans there are two different types: HERV-Fc1 and HERV-Fc2. HERV-Fc1 has a full provirus, but only the *env* gene has an ORF. There are five copies of HERV-Fc2 with mostly present proviruses, of the five, one HERV-Fc2Δenv has an *env* ORF. However, HERV-Fc2Δenv *env* ORF is truncated before the MSD due to a deletion of the 3' end of the provirus [74]. In addition to humans they found ERV-Fc with an intact *env* ORF in baboons (babERV-Fc2). Based on LTR divergence it was estimated the HERV-Fc2 integrated into the primate genome

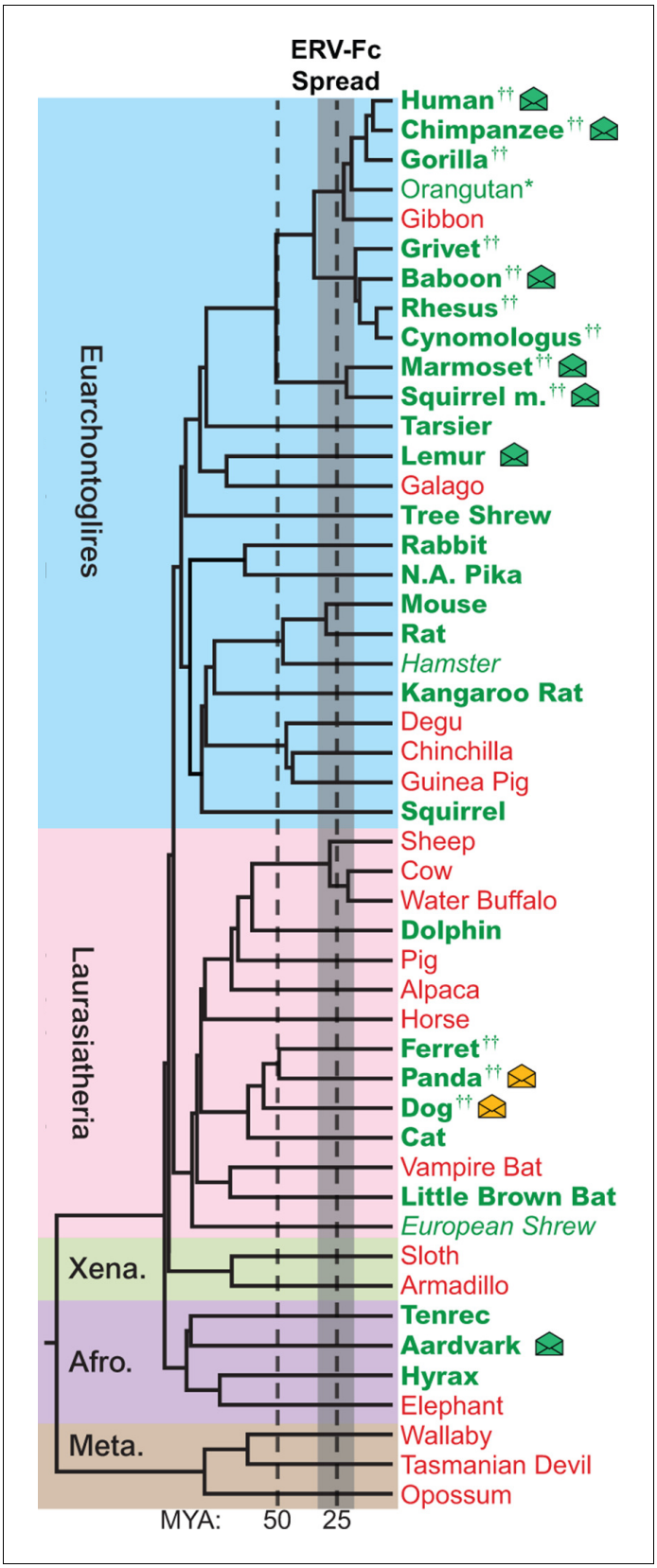


Figure 1.5: ERV-Fc in Eutherian mammal genomes. Species that ERV-Fc were not found in are in red, species with ERV-Fc in their genome are in green. If one or more ERV-Fc gene could be reconstructed the name is in bold, italics indicate fragments were identified but they could not be reconstructed. * = only a solo LTR was found, † † = the species has two genetically distinct ERV-Fc lineages in is genome. Envelope icons indicated an *env* ORF(s) was found in that species, green ones indicate they are classical ERV-Fc *envs* and orange icons indicate the *env* originated from a crossover event. (Figure adapted from Diehl *et al.* 2016.)

20 to 32 million years ago, and that babERV-Fc2 is a recent integration due to the identical LTRs [74]. Heidmann was also able to confirm through polymerase chain reaction (PCR) the presence of HERV-Fc1 and two of the HERV-Fc2s at the same locus in chimpanzees and gorillas [74]. A large real-time quantitative PCR screen of HERV *env* ORF transcript levels showed low levels of HERV-Fc1 transcripts in human skin, testis and trachea tissue, and HERV-Fc2Δenv transcripts in skin and testis tissue [76].

Using Basic local alignment search tool nucleotide (BLASTn), we previously screened 50 mammalian genomes searching for ERV-Fc loci. ERV-Fc loci were found in 28 of those genomes, including previously characterized dog ERV-Fcs (cfERV-Fc) (Fig. 1.5) [77], [78]. Additionally, *env* ORFs were found in marmosets, squirrel monkeys, lemurs, aardvarks, pandas and dogs; adding to the list of ERV-Fc *env* ORFs previously discovered in humans, chimpanzees and baboons (Fig. 1.5) [77]. By looking at LTR divergence, it was determined that ERV-Fc was spreading and infecting a wide host of mammals between ~33 million to ~15 million years ago [77]. Phylogenetic analysis determined that at least 26 cross-species transmission events gave rise to the identified ERV-Fc insertions, though as ERV-Fc spread exogenously the number was probably greater as the ERV-Fc "fossil" record can only give a narrow picture of how the virus actually spread [77]. During the phylogenic analysis, it was also discovered that ERV-Fc has experienced several recombination events. At least one recombination event has led to an ERV-Fc in carnivores to acquire an *env* that aligned more with HERV-W than ERV-Fc *envs* (Fig. 1.5). The recombinant ERV

then infected dogs, pandas and ferrets, resulting in the *env* ORFs in dogs (cfERV-Fc1(a)-env) and pandas (ameERV-Fc1-env) (Fig. 1.5) [77].

Gammaretrovirus Envelope Glycoproteins

This thesis describes the reconstitution and functional characterization of the Env proteins of the ancient lineage of gammaretrovirus that gave rise to ERV-Fc elements in mammalian genomes. Gammaretrovirus Envs have the canonical signal peptide at the beginning of their sequence as well as a furin cleavage site between the SU and TM (Fig. 1.3). The gammaretrovirus Env also has several distinctive features. The SU and TM domains are covalently associated; a CXXC isomerization motif in the SU and a CX6CC motif in the TM are responsible for this interaction (Fig. 1.3). Additionally, in the TM there is an immunosuppressive domain (ISD) directly before the CX6CC motif (Fig. 1.3). The fusion peptide in gamma Envs is typically found in TM immediately after the furin cleavage site (Fig. 1.3).

Also, unique to gamma-like Envs is the R-peptide, a short sequence at the end of the cytoplasmic tail that is cleaved off by the viral protease after virion assembly to activate the Env's fusogenic ability (Fig. 1.3). The R-peptide has been studied in both MLV and MPMV. The MLV R-peptide was first discovered when two forms of the TM subunit were observed, one with a slightly lower molecular weight originally termed p15(E) and p12(E) [18]. Through use of a sequence specific antibody the cleavage product was confirmed. It was also found that it (the R-peptide) was cleaved off during the final viral maturation stage [19], [20]. Reversed-phase high-pressure liquid chromatography was used

to separate MLV protein products and the R-peptide was isolated and sequenced. It was determined that the R-peptide was the last 16 amino acids of the TM subunit and the cleavage occurred between a leucine (L) and a valine (V) [21]. Expression of truncation mutants of the MLV Env resulted in cell-cell fusion when only the R-peptide was missing. This cell-cell fusion did not occur when full length MLV Env was expressed. This indicated that cleaving the R-peptide off the Env activated the fusion ability of the Env [22], [23].

Similar to MLV, MPMV's R-peptide was discovered when TM was observed to have two different molecular weights. During a pulse-chase experiment it was noted that within the cell the TM was in a gp22 form. Once the virions budded the gp22 form began to decrease while a gp20 band increased until it was the major form of TM, indicating about 16 amino acids were being cleaved off with a probable cleavage site between a tyrosine (Y) and a histidine (H) [79]. Because the cleavage happens in virion and not the host cell, it was thought the viral protease must be responsible for the processing. To confirm the hypothesis the viral protease was mutated and a protease inhibitor was used. Applying either method the Env was still incorporated into virions; however, processing between gp22 and gp20 was impaired as was infection [79], [80]. As with MLV, when truncation mutants were made in the MPMV TM cytoplasmic tail cell-cell fusion was greatly enhanced [81].

Further confirmation of R-peptides in gammaretroviral Envs and the effect cleaving off the R-peptide has on activating fusion was done with three viruses with gamma Envs: gibbon ape leukemia virus (GaLV), spleen necrosis virus

(SNV), and porcine endogenous retrovirus (PERV) [82]. They all have the same L-V motif in their cytoplasmic tail as MLV, truncations were made to this site and the mutant Envs were expressed in cells. Cell-cell fusion was observed for the truncated Envs [82]. In addition to being important for fusion of the virion to a new host cell, the R-peptide may be involved in targeting the Env for incorporation into virions. It was shown with MPMV and MLV that truncating the cytoplasmic tail to the R-peptide or introducing mutations in certain areas of the R-peptide led to lower levels of Env incorporation into virions [81], [83]. The gammaretroviral R-peptide therefore plays a role both in incorporation into virions and in the activation of the Env's fusogenic ability.

Host Co-option of ERVs (Exaptation)

The work in this thesis also reflects the potential evolutionary cooption of ERV Env for host functions. As more ERVs have been discovered with genome sequencing, the question of why some have maintained ORFs has been investigated, leading to the discovery that multiple ERVs have been exapted by the host for a beneficial function. The function of these ERVs tend to fall into two categories: use of the Envs fusogenic ability or use as an anti-viral factor (Fig. 1.6 and 1.7) [34].

Syncytins

The use of an ERV *env* for its fusogenic ability is exemplified by the *syncytin* genes. *Syncytins* are ERV genes involved in the formation of the syncytiotrophoblast layer in the placenta, a multinucleated syncytium (Fig. 1.6) [84]. There are several characteristics an ERV Env is required to have before it

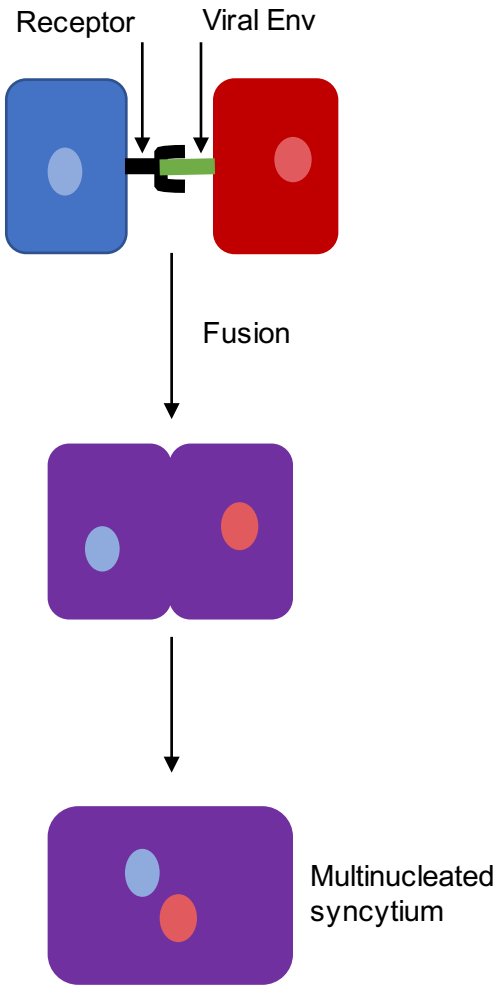


Figure 1.6: Formation of a syncytium. Depiction of how a multinucleated syncytia forms, when one cell expresses a viral Env, the Env can bind a receptor on a neighboring cell (top) and trigger fusion (middle) resulting in the two cells fusing together into one cell with multiple nuclei (bottom).

can be considered a *syncytin*: first, it needs to be expressed in the placenta; second, it must have maintained its cell-cell fusion ability (Fig. 1.6); finally it should be highly conserved among related taxa [85]. Through the use of these criteria, multiple *syncytins* have been found in a wide swath of mammals including: primates, carnivores, ruminants, rodents, marsupials, tenrecidae and leptoridae [86], [87]. Most of these *syncytins* appear to have emerged independently as early as 85 million years ago. However, the first placental mammals emerged roughly 150 million years ago, indicating there may have been an original captured ERV Env *syncytin* that over time has been replaced by other ERV Envs in different species [87].In humans there appear to be two main *syncytins,* a HERV-W *env* and an HERV-FDR *env,* known as *syncytin-1* and *syncytin-2,* respectively [84], [87]–[89]. Both are highly expressed in the placenta cells involved with the formation of the syncytiotrophoblast layer [84], [87], [89]–[91]. *Syncytin-1* is highly conserved in Hominids, having entered the genome before the split with Old World monkeys, making it roughly 30 million years old [92]. It uses the RD-114 and D-type retrovirus (RDR) supergroup receptor ASCT2, a neutral amino acid transporter [90]. *Syncytin-2* is conserved in all primates except prosimians and is estimated to be 45 million years old [93]. Both of these ERV Envs can mediate fusion of cells in culture; silencing of either gene leads to an impairment of cell-cell fusion [87], [94]–[98]. In addition to their fusogenic ability there is some evidence that Envs have an immunosuppressive function because of the ISD [86]. In a tumor rejection assay in mice it was found that *syncytin-2,* but not *syncytin-1,* is immunosuppressive [99], [100]. This

suggests a potential role for *syncytins* in immunosuppression to help prevent fetal rejection by the mother.

While it is almost impossible to confirm whether the human *syncytins* are involved in the formation of the placenta, proof of concept was tested in mice. After the sequencing of the mouse genome two potential *syncytin* genes were found, *syncytin-A* and *syncytin-B.* Even though the mice *syncytins* are genetically distinct from the human *syncytins* they have some of the same characteristics: they are expressed specifically in the placenta, are fusogenic and are highly conserved from when they were endogenized approximately 25 million years ago [101]. Similar to their human counterparts, *syncytin-B* is immunosuppressive while *syncytin-A* is not [100]. To test the actual importance of *syncytins* in pregnancy, knockout mice were created. When *syncytin-A* was knocked out the embryos died at midgestation; there was significant placental architectural defects and an accumulation of unfused cells [102]. When *syncytin-B* was knocked out the phenotype was not as severe, the animals were still viable, but there was growth retardation and fewer offspring [103]. A double knockout of both *syncytin* genes demonstrated a more severe phenotype than the *syncytin-A* knockout with the embryos dying earlier than in the single knockout, indicating that both *syncytins* are required for proper placenta formation [103].

ERVs coopted for an Antiviral function

In addition to ERVs being coopted for their fusogenic ability they have been exapted to function as antiviral factors (Fig. 1.7). One of the first to be

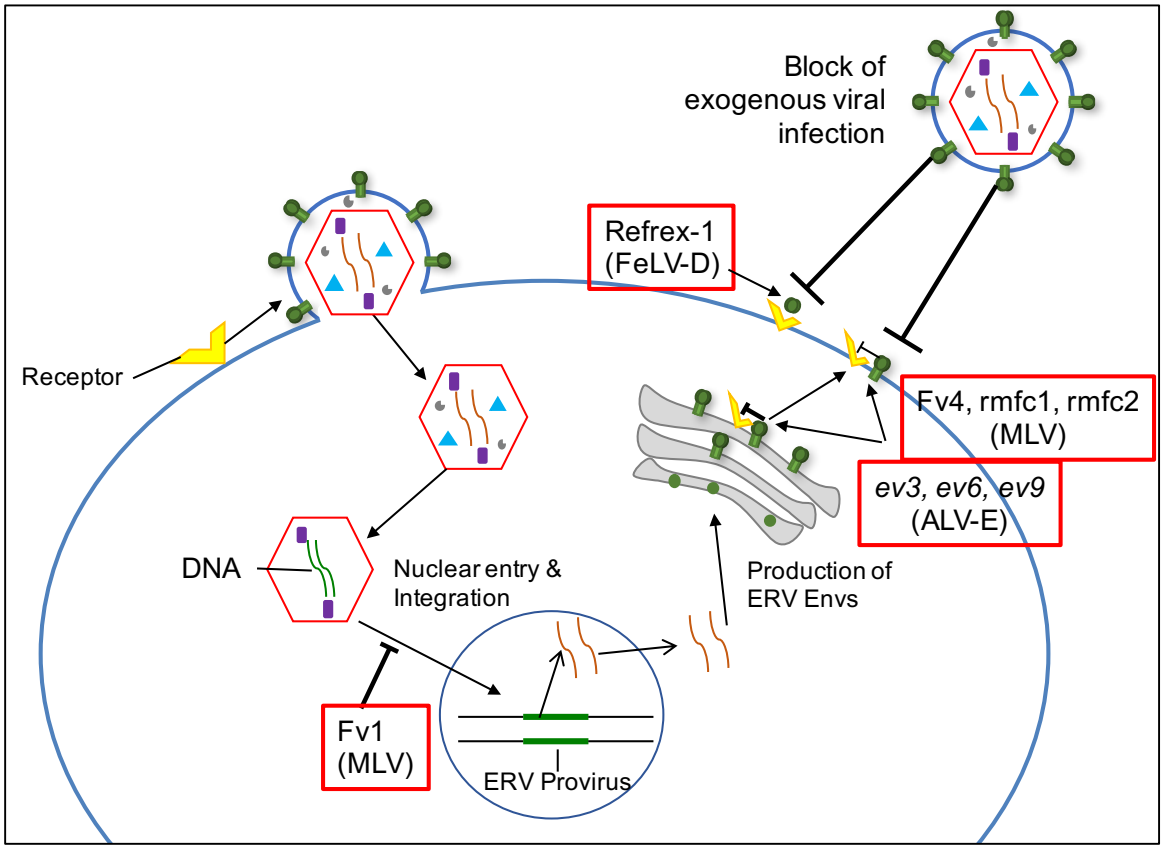


Figure 1.7: Antiviral ERVs. Representation of where ERV proteins act as antiviral agents. Fv1 acts early in the viral life cycle before integration potentially binding to the viral capsid. Fv4, rmfc1, rmfc2, *ev3, ev6, ev9* and Refrex-1 are all ERV Envs that use receptor interference, blocking the receptors from use by an exogenous virus. This may occur in the ER at the cell surface or in the case of Refrex-1 in the extracellular space after being secreted. The virus these factors block are in parentheses bellow the factor (red boxes).

discovered was friend virus susceptibility factor 1 (*Fv1*) in mice, which protects mice from infection by exogenous MLV (Fig. 1.7). First characterized in the 1970s it was not originally known to be of ERV origin, but it was observed that mice with the *Fv-1* locus were resistant to MLV [104]. There are two different loci of *Fv-1* that lead to the identification of three variants of MLV. *Fv-1*^{n} found in NIH/Swiss mice confers resistance to N-tropic MLV, *Fv-1*^{b} found in Balb/c mice confers resistance to B-tropic MLV, but not visa-versa. Heterozygous *Fv-1*^{n}*/Fv-1*^{b} mice are resistant to both N- and B-tropic MLV. Additionally, there is a NB-tropic MLV that can avoid being blocked by either locus of *Fv-1* [104]–[106]. It was not until the 1990s when the *Fv-1* locus was cloned that it was discovered that it encodes a Gag-like protein similar to the endogenous families HERV-L and murine ERV-L (MuERV-L) [107] . Recent studies have found orthologous *Fv-1* in rodents, outside of just mice, suggesting an endogenization event at least 45 million years ago [108]. Cells stably expressing *Fv-1* and challenged with EIAV and feline foamy virus (FFV) showed a reduced infection rate, indicating *Fv-1* have the potential to protect against a wide range of viruses [109]. The mechanism of *Fv-1* restriction is not clear, but it is known that it acts early in the viral life cycle before integration, and it is thought to bind to the entering capsid core, similar to TRIM5α, but how infection is prevented from continuing is not known (Fig. 1.7) [105], [110]–[112].

Receptor interference is another antiviral method that ERVs have been coopted for, which involves the *env* gene (Fig. 1.7). When expressed they may bind to a receptor blocking use by another viral Env, consequently protecting that

host from infection from an exogenous virus that uses the same receptor (Fig. 1.7). There are multiple recorded instances of this occurring: in mice, there is friend virus susceptibility factor 4 (*Fv-4*), *rmfc1* and *rmfc2,* in chickens endogenous ALVs *ev3, ev6* and *ev9,* and in cats Refrex-1 (Fig. 1.7) [113]–[117].

The first ERV Envs discovered to conferred receptor interference were endogenous ALVs. ALV has several subgroups based on their Envs: A, B, C, D and E. Subgroup E originates from recombination events with endogenous ALV, which can then infect and spread in a chicken population [113]. It was observed after doing crosses of chickens, that some chickens had low susceptibility to infection by ALV-E, but were susceptible to ALV-B and C [113]. When the endogenous viral loci were observed, it was found that chickens with *ev3, ev6* and *ev9* loci were resistant to ALV-E and chickens that did not have these loci were susceptible to infection by ALV-E. Chickens with only *ev3* were intermediately protected and those with *ev6* and *ev9* were highly protected. Cultured cells expressing these Envs were also able to protect against infection [113]. This suggests that *ev3, ev6* and *ev9* have been evolutionarily maintained to protect chickens from the horizontal spread of ALV-E, and the likely mode of protection is by blocking the receptor analogous to receptor interference (Fig. 1.7).

Fv-4 was thought to be a gene in mice that made them resistant to Friend-MLV (F-MLV) or ecotropic MLV [115]. An Asian mouse, strain G, was found to be resistant to N- B- and NB-MLV indicating a different restriction factor than *Fv-1,* it was also identified in a wild mouse population in southern California around

Lake Casitas [115], [118]–[120]. During the investigation into *Fv-4* it was discovered that it resembled a MLV Env and presumably acted through receptor interference (Fig. 1.7) [121], [122]. An *env* specific probe was used to identify *Fv-4*s sequence and to confirm it was similar to exogenous MLV *env* [123]. When fully sequenced it was found that *Fv-4* is part of a defective provirus, with the only remaining pieces belonging to the 3' end of *pol*, the full *env* and the 3' LTR [124]. While the *Fv-4* is able to block the F-MLV receptor it is incapable of infecting cells, when virus was pseudotyped with full length *Fv-4* it was noninfectious [125]. The defect in ability of *Fv-4* to mediate entry was later mapped to the fusion peptide, where there is an arginine (R) in place of a highly-conserved glycine (G), this point mutant reduced the ability of *Fv-4* to mediate fusion [126]. Fusion, therefore, is not a necessary function for Envs to confer receptor interference.

In mice, there are two other ERV *envs* that confer resistance to exogenous virus, they are *rmfc1* and *rmfc2*. *Rmfc1* was first discovered in 1983, when it was observed that DBA/2 mice were resistant to several forms of MLV including a recombinant MLV called mink cell focus-forming (MFC) virus. The resistance was found to be separate from *Fv-1* restriction and thought to involve restriction of the incoming Env [116]. Using type-specific antibodies it was confirmed that *rmfc* was an endogenous ERV *env* gene, and it restricts through receptor interface [127]. This was later proved when the full length provirus was found after doing mouse crosses with mice without *rmfc1*; only the *env* of the provirus was intact, the *gag* and *pol* genes were defective due to deletions [128].

Rmfc2 blocks infection by polytrophic MLV. Through genetic crosses *rmfc2* was discovered to be an ERV *env* and was associated with a full provirus that had a large deletion in integrase [117], [129]. In mice, this brings the total of ERV *env* antiviral genes up to three; all are distinctly different from each other but act in a similar manner (Fig. 1.7).

There are several endogenous retroviruses in cats; one of the families is endogenous gammaretrovirus of domestic cats (ERV-DCs). There are at least 19 insertions of the ERV lineage in cats [130]. These ERVs are polymorphic (not fixed) except ERV-DC7 and ERV-DC16. LTR divergence of ERV-DC7 indicates it was endogenized 2.8 million years ago. ERV-DC10 and EV-DC18 are completely intact and can still make replication-competent virus that has a broad infectious tropism [130]. A recombination event between FeLV and ERV-DC *envs* resulted in a new subgroup, FeLV-D, though it is not clear whether this is currently a spreading virus capable of producing infectious virus [130]. While determining the receptor interference groups of FeLV-D and the ERV-DCs, it was discovered that the supernatant from feline 3201 cells blocked infection from FeLV-D and genotype I ERV-DCs. The phenotype was additionally observed with supernatant from other cat cells, but not other species such as human or dog [114]. It was hypothesized that there was a secreted restriction factor in supernatant from cat cells, later named restriction for feline retrovirus X (Refrex-1). cDNA produced from 3201cells mRNA indicated that Refrex-1 came from two loci, the ERV-DC7 and ERV-DC16 *envs* [114]. Both of these Envs have an early stop codon truncating them in the SU after the hypothetical receptor binding

domain. They are expressed in cat tissues and supernatant from cells transfected with ERV-DC7 or ERV-DC16 inhibit infect by FeLV-D, indicating that Refrex-1 is expressed from both of the loci [114]. Refrex-1 acts through receptor interference, likely after it has been secreted from the cell, and may have been maintained to protect against the reemergence of ERV-DCs, contributing to their endogenization (Fig. 1.7) [114]. Interestingly even when the full length ancestral Refrex-1 was reconstructed furin was unable to cleave the *env* into the SU and TM subunits, this defect was due to a mutation upstream of the furin cleavage site [131]. The mutation and the early stop may have contributed to ERV-DC7 and ERV-DC16s exaptation and use as an antiviral factor.

Recently a study was done on HERV-T that suggests it might have had an antiviral function. HERV-T entered the primate germline ~43 to 32 million years ago with the most recent fixed integrations occurring about 11 million years ago [132]–[134]. A reconstructed ancestral HERV-T *env* (ancHTenv) was able to infect a wide range of mammalian cells and was found to use human monocarboxylate transporter 1 (hMCT1) as a receptor [134]. Interestingly the human genome has a copy of HERV-T with an almost complete *env* ORF, lacking only five amino acids from the C-terminus [76]. Similar to Refrex-1 and *Fv-4* this HERV-T Env (hsaHTenv) is defective in several ways [114], [125], [126], [131]. The HERV-T Env is not correctly processed and or incorporated into virions, therefore, it is non-infectious [134]. This is due in part to mutations in the furin cleavage site; however, comparable to Refrex-1 when the cleavage site was reconstructed, furin cleavage was not restored, indicating there may be

additional deactivating mutations [131], [134]. Even with these defects, when hsaHTenv was expressed in cells with the receptor and then challenged with virus containing ancHTenv it was able to block infection, acting in an antiviral fashion. It appears to achieve this block by reducing the amount of hMCT1 at the cell surface, as seen in a depletion of hMCT1 in western blots and tagged hMCT1 from cell surfaces [134]. The hsaHTenv provirus was inserted into the germline around 13 to 19 million years ago, and during that time has been under selective pressure to maintain the *env* ORF [134]. All of these combined suggest that hsaHTenv was preserved for an antiviral function and may have contributed to the elimination of the exogenous form of HERV-T [134].

This thesis describes functional characterization of the *env* ORFs of the ERV-Fc family. We hypothesize that these ORFs have been maintained for a cellular function, similar to *Fv-4, rmfc-1* and *rmfc-2* in mice. Functional characterizations of the Envs provides insight into exaptation of ERVs and a possible function that the ERV-Fc Envs may have performed.

Chapter 2: Materials and Methods

Constructs

Codon optimized sequences for expression in human cells of: human, chimpanzee, bonobo and gorilla-Fc1 consensus (conERV-Fc1), human-Fc2Δenv (HERV-Fc2Δenv), baboon-Fc2 (babERV-Fc2), marmoset-Fc3-1 (cjaERV-Fc3-1), squirrel monkey-Fc3-1 (sboERV-Fc3-1), grey mouse lemur #1 (gmlERV-Fc-#1), grey mouse lemur #2 (gmlERV-Fc-#2), aardvark-Fc1 (oafERV-Fc1), dog-Fc1(a) (cfERV-Fc1(a)) and panda-Fc1 (ameERV-Fc1) were synthesized (GenArt-Thermo Fisher Scientific). Sequences were then cloned into pcDNA3.1+ using the restriction enzymes EcoRI-HF and NheI-HF (NEB). A non-codon optimized sequence of babERV-Fc2 was also synthesized and cloned into pcDNA3.1+. Codon optimized sequence of HERV-FcΔenv with a membrane spanning domain and the MLV cytoplasmic tail (HERV-Fc2Δenv-MLVct) were synthesized and cloned into pcDNA3.1+. All synthesized sequences contain a c-terminal Avi tag.

Around-the-horn PCR was used to modify the furin cleavage site of babERV-Fc2, HERV-Fc2Δenv and conERV-Fc1 (Table 2.1) [135]. Primers were generated to flank the site of interest, resulting in babERV-Fc2-cl (IQKQ to RQKR), HERV-Fc2Δenv-MLVct-babSP-cl (KSKR to RQKR), conERV-Fc1-375+ASQLS, conERV-Fc1-H340T, conERV-Fc1-F335G, conERV-Fc1-F335G, 375+ASQLS, conERV-Fc1-375+ASQLS, P387L, conERV-Fc1-F335G, P387L and conERV-Fc1-P387L. PCR reactions were then digested with DpnI (NEB) for 1.5 hours at 37°C and then ligated back together using the Promega T4 rapid ligase. Cytoplasmic tail truncation mutants of babERV-Fc2-cl were constructed

Primer Name	Primer
ERV-Fc-F	AAGCTGGCTAGCGCCACCATG
ERV-Fc-R	CTCTCTGAATTCTCATTCGTGCCACTC
Bab-ERV-Fc CS F	CGCCAGAAACGGGCCGTGTTCCTGCCTCTG
Bab-ERV-Fc CS R	GTTGTTGCTGCTGCTCAGCTG
Bab Trunc 2	ctctctGAATTCttattcatgccattcaattttctgcgcttcaaaaatatcgttcaggccagctcctgctc c CACTCTGGTGATTTCCTGCACTCTGG
Bab Trunc 3	ctctctGAATTCttattcatgccattcaattttctgcgcttcaaaaatatcgttcaggccagctcctgctc c GTAGGGATGCAGCAGCATCTGG
Bab Trunc 4	ctctctGAATTCttattcatgccattcaattttctgcgcttcaaaaatatcgttcaggccagctcctgctc c GGGGTCGCTGGTGGGC
Bab-MLV-Ct	ctctctGAATTCttattcatgccattcaattttctgcgcttcaaaaatatcgttcaggccagctcctgctc cTTGTGGCTCGTATTCTAGTGGTTTTAGCTGGTGGTATTGTTGAGTCAGGACTAAAGCCTGGACTACTGAG ATCCTGTCTTTCAGGAACTTAATCAGGATAGGGGCCAG
Aard-MLVct	ctctctGAATTCttattcatgccattcaattttctgcgcttcaaaaatatcgttcaggccagctcctgctc cTTGTGGCTCGTATTCTAGTGGTTTTAGCTGGTGGTATTGTTGAGTCAGGACTAAAGCCTGGACTACTGAG ATCCTGTCTTTCAGGAACTTCAGCAGGCAAG
Lem#2-MLVct	ctctctGAATTCttattcatgccattcaattttctgcgcttcaaaaatatcgttcaggccagctcctgctc cTTGTGGCTCGTATTCTAGTGGTTTTAGCTGGTGGTATTGTTGAGTCAGGACTAAAGCCTGGACTACTGAG ATCCTGTCTTTCAGGAACTGCAGCAGGC
Dog-MLVct	ctctctGAATTCttattcatgccattcaattttctgcgcttcaaaaatatcgttcaggccagctcctgctc cTTGTGGCTCGTATTCTAGTGGTTTTAGCTGGTGGTATTGTTGAGTCAGGACTAAAGCCTGGACTACTGAG ATCCTGTCTTTTAAACATCGAAACAAGCAAGGGGCGAC
Fc2-Fc1SP B F	AAGCTGGCTAGCGCCACCATGGCCAGACCTAGCCCTCTGTGTCTCCTGCTGCTGCTGACCCTGCTGCCCCC TATCGTGCCCAGCAATAGCCTGCTGACC AACAGCCCCTGCGACAGAC
Fc2-BabSP B F	AAGCTGGCTAGCGCCACCATGATCAGCGCCGTGCTGAACCTGCCTAGCACCCCTCTGCTGCCCCTGCTGTG GTTCACCCTGATCATCCCTGCCAGCCTGACCAACCCCAAGTTCGTGAACAGCCCCTGCGACAGAC
Fc2-BabSP NB F	AAGCTGGCTAGCGCCACCATGATCAGCGCCGTGCTGAACCTGCCTAGCACCCCTCTGCTGCCCCTGCTGTG GTTCACCCTGATCATCCCTGCCAGCCTGACCAACCCCAAGTTCGTGGGCTTTAGCAGCCTGACCGAG
Fc2-StoQ F	CAGAAGCGGGCCATCTTCCTG
Fc2-KtoR R	CCGGTTCTGGGGTCTGGGCTC
Fc1 P to L F	TGCCTCTCGTGATCGGCGTG
Fc1 P to L R	GGAACACGGCTCTTTTCTGCC
Fc1 5AA gap F	GCCTCCCAGCTGAGCAATCCCCCCATGCGGC
Fc1 5AA gap R	CACGAGGGAGGACAGCTCG
Fc1 F to G F	GGCACCCTGACCAAGCACCTG
Fc1 F to G R	ATTGCACCAGAAGTATCCGCC
Fc1 H to T F	ACCCTGAACATCAGCAGCAACAATAC
Fc1 H to T R	CTTGGTCAGGGTGAAATTGCAC

Table 2.1: Primers used to make constructs.

using PCR and primers flanking the desired sequence, reverse primers contained the Avi tag (Table 2.1). Four truncation mutants were made: babERV-Fc2-cl-Δ4AA, babERV-Fc2-cl-Δ12AA, babERV-Fc2-cl-Δ22AA and babERV-Fc2-cl-Δ29AA. All PCR products were gel purified and cloned into pcDNA3.1+ using the restriction enzymes EcoRI-HF and NheI-HF.

Chimeric env's containing the MLV cytoplasmic tail (MLVct), the babERV-Fc2 signal peptide (babSP) or the conERV-Fc1 signal peptide (Fc1SP) were constructed using PCR and primers that flanked the desired sequence and contained either the MLVct+Avi tag or the signal peptides (Table 2.1). Chimeras generated were: babERV-Fc2-cl-MLVct, gmlERV-Fc-#2-MLVct, oafERV-Fc1-MLVct, cfERV-Fc1(a)-MLVct, HERV-Fc2Δenv-babSP, HERV-Fc2Δenv-babSPintr, HERV-Fc2Δenv-MLVct-babSP, HERV-Fc2Δenv-MLVct-babSPintr, and HERV-Fc2Δenv-MLVct-Fc1SP. All PCR products were gel purified and cloned into pcDNA3.1+ using the restriction enzymes EcoRI-HF and NheI-HF.

Cell lines and cell culture

Cell lines were obtained from the American Type Culture Collection (ATCC). Cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM): HEK293T/17, MDCK, Hos, LLC-MK2, CRFK, DF-1, or Eagle's Minimum Essential Medium (EMEM): HT1080, Vero or F-12 medium, A549. Media was supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine (200mM), 1% Penicillin-Streptomycin (10,000 IU- 10,000 µg/ml) and 2.5% HEPES (1M). Cells were incubated at 37°C except DF-1, which were incubated at 39°C.

Western blots

293T/17 cells seeded in 6-well plates were transfected with 1μg of plasmids containing the ERV-Fc-*envs* using GenJet (SignaGen). 48 hours after transfection cell lysates were collected by directly lysing cell in the wells with IP lysis buffer (50mM Tris-HCl, 150mM NaCl, 1mM EDTA, 1% Triton x-100, 5% glycerol, pH 7.5). Lysed cells were centrifuged at 4°C and 14,000 rpm and the supernatant was saved. Supernatants were then boiled with 2X laemmli buffer and loaded onto a 12% SDS-PAGE gel for electrophoresis. Proteins were transferred to a PVDF membrane, the membrane was blocked with 1X-PBS-Tween80-5% milk. Membranes were then probed with monoclonal mouse anti-Avi (Avidity) then goat anti-mouse-HRP conjugated antibody (Thermo Scientific). Blots were developed with Amersham ECL Select Western Blotting Detection Reagent (GE Healthcare) and detected using a ChemiDoc MP imaging system (Biorad). Blots were then striped with Restor™ Western Blot Striping Buffer (Thermo Scientific), re-blocked with 5% milk then probed with anti-βactin-HRP and imaged again.

Pseudotyping and Infectivity

MLV particles were pseudotyped by transfecting 293T/17 cells with pLXIN-GFP, pCIGB and a plasmid encoding one of the following glycoproteins: vesicular stomatitis virus glycoprotein (VSVG), babERV-Fc2-cl-MLVct, gmlERV-Fc-#2-MLVct, oafERV-Fc1-MLVct, cfERV-Fc1(a)-MLVct**.** The plasmids were transfected at a ratio of 3:2:1 respectively using GenJet (SignaGen® Laboratories), total DNA: 2μg (6 well plates), 4μg (T25s) or 8μg (T75s). MPMV

particles were pseudotyped by transfecting 293T/17 cells with pSARM4 and pTMO at a ratio of 1:1 using GenJet, and the same DNA totals listed above.

48 hours after transfection supernatant was centrifuged at 5000rmp for 5 minutes and either added directly to cells for infection or viral particles were concentrated using Centriprep® 50k filter devices (Merck Millipore). 12 well plates of cells at ~50% confluence were used for infection, cells infected include: 293T/17, Hos, HT1080, A549, Vero, LLC-MK2, CRFK, MDCK, and Df-1. 250µl of supernatant containing MLV-NoEnv and MLV-VSVG pseudotyped particles was added to cells, 900µl of supernatant of other pseudotyped particles was added to cells for infection. 4 hours after infection 500µl of fresh D10 was added to cells. 48 to 96 hours after infection, cells were imaged using an EVOS microscope then harvested for flow cytometry.

Superinfection Interference assay

293T/17 cells were seeded in 12 well plates for a confluence of ~40%, then were transfected with 750ng of: pcDNA3.1+ empty vector, VSVG, SIVgp160, conERV-Fc1, HERV-Fc2Δenv, HERV-Fc2Δenv-MLVct-babSP, babERV-Fc2, babERV-Fc2-cl-MLVct, cjaERV-Fc3-1, oafERV-Fc1, gmlERV-Fc-#2 or sboERV-Fc3-1 after 24 hours. Pseudotyped MLV virus was generated as before, with either no Env, VSVG or babERV-Fc2-cl-MLVct for Envs. 48 hours after transfection, viral containing supernatant was collected and used to challenge cells transfected with Envs, amounts of viral supernatant used were the same as above. 48 hours after infection cells were imaged using an EVOS microscope then harvested for flow cytometry.

Flow/FACS

Cells harvested for flow were fixed with 2% paraformaldehyde in 1xPBS. The number of GFP positive cells were counted by FACS using BD FACSAria cell sorter (BD Biosciences). The data was analyzed with FlowJo (version 8.7.3, FlowJo LLC), % of GFP positive cells was calculated after gating on live cells.

Bioinformatics and Phylogenetics

For BLAST searches, we used NCBI BLASTn and BLASTp functions. All sequence alignments and annotations were done in Geneious (version 11.0.5 or earlier, https://www.geneious.com), using the multiple alignment tool and annotation features [136]. Phylogenetic and sequence diversity analysis were also conducted using the Geneious tree builder tool (neighbor-joining) and the distances tool after constructing an alignment. To identify putative signal peptides in ERV-Fc Envs, we used the predicted amino acid sequences and the SignalP4.1 server (http://www.cbs.dtu.dk/services/SignalP/) [137].

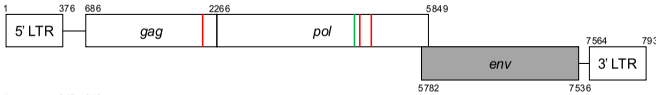
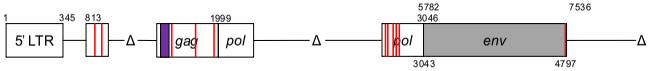
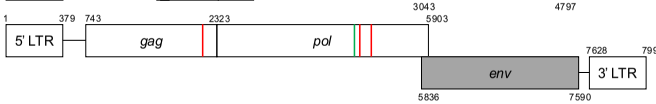
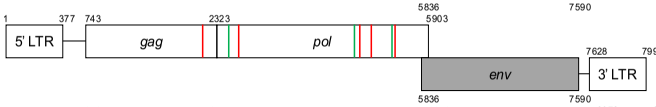
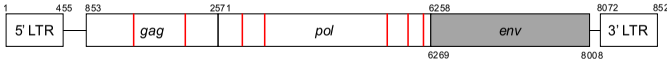
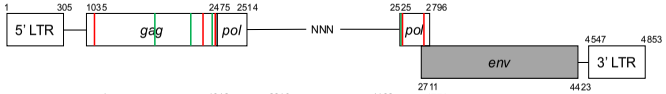
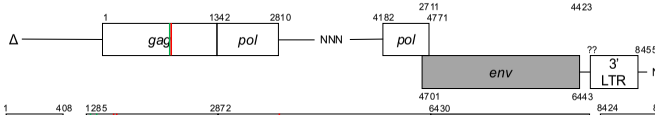
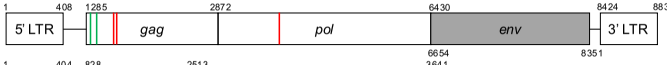
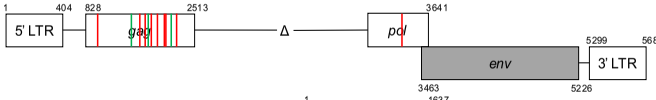
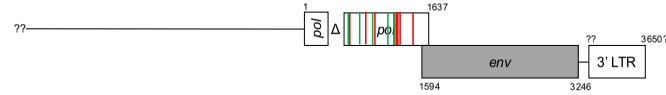
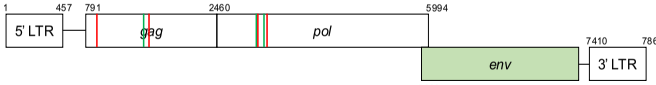
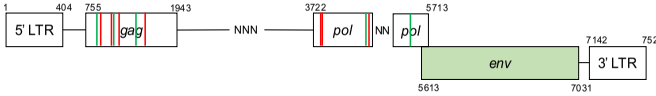
Chapter 3: Results

ERV-Fc env ORF proviruses are disrupted

ERV-Fc-*env* ORFs in the human and baboon genomes were previously described by the Heidmann lab [74], [75]. Both HERV-Fc1 and babERV-Fc2 are full-length proviruses with 5'-3' LTR identities of 92.8% and 100% respectively. HERV-FcΔenv is truncated just upstream of the membrane spanning domain (MSD) in *env* and is missing the 5' LTR (Fig. 3.1) [74]. More recently, we performed an extensive survey of 50 mammalian genomes and discovered ERV-Fc *env* with intact ORFs in the genomes of chimpanzee (homERV-Fc1), bonobo (ppaERV-Fc1), marmoset (cjaERV-Fc3-1), squirrel monkey (sboERV-Fc3-1), grey mouse lemur (gmlERV-Fc-#1 and gmlERV-Fc-#2), aardvark (oafERV-Fc1), dog (cfERV-Fc1(a)) and panda (ameERV-Fc1). Phylogenetic analysis revealed that the panda and dog *envs* originated from a recombination event, while the afore mentioned are "classical" ERV-Fc *envs* (Fig. 3.1) [77].

To reconstruct the proviral sequences in which the newly discovered ERV-Fc *env* ORFs are found, I used BLAST and extracted flanking sequences 10kb upstream and 10kb downstream of each *env* gene, and analyzed these for presence of viral sequences and ORFs related to *gag, pro* and *pol* genes of retroviruses. I found that none of the proviruses have maintained intact *gag* or *pol* ORFs, although remnants of the genes were present (Fig. 3.1). The chimpanzee and bonobo full-length proviruses are at the same locus as HERV-Fc1 (i.e. these are othologous loci). The HERV-Fc1 provirus is also present in gorillas, but contains an early stop codon in the *env* gene ORF. The other full length proviruses include gmlERV-Fc-#1 and cfERV-Fc1(a). Genome sequence

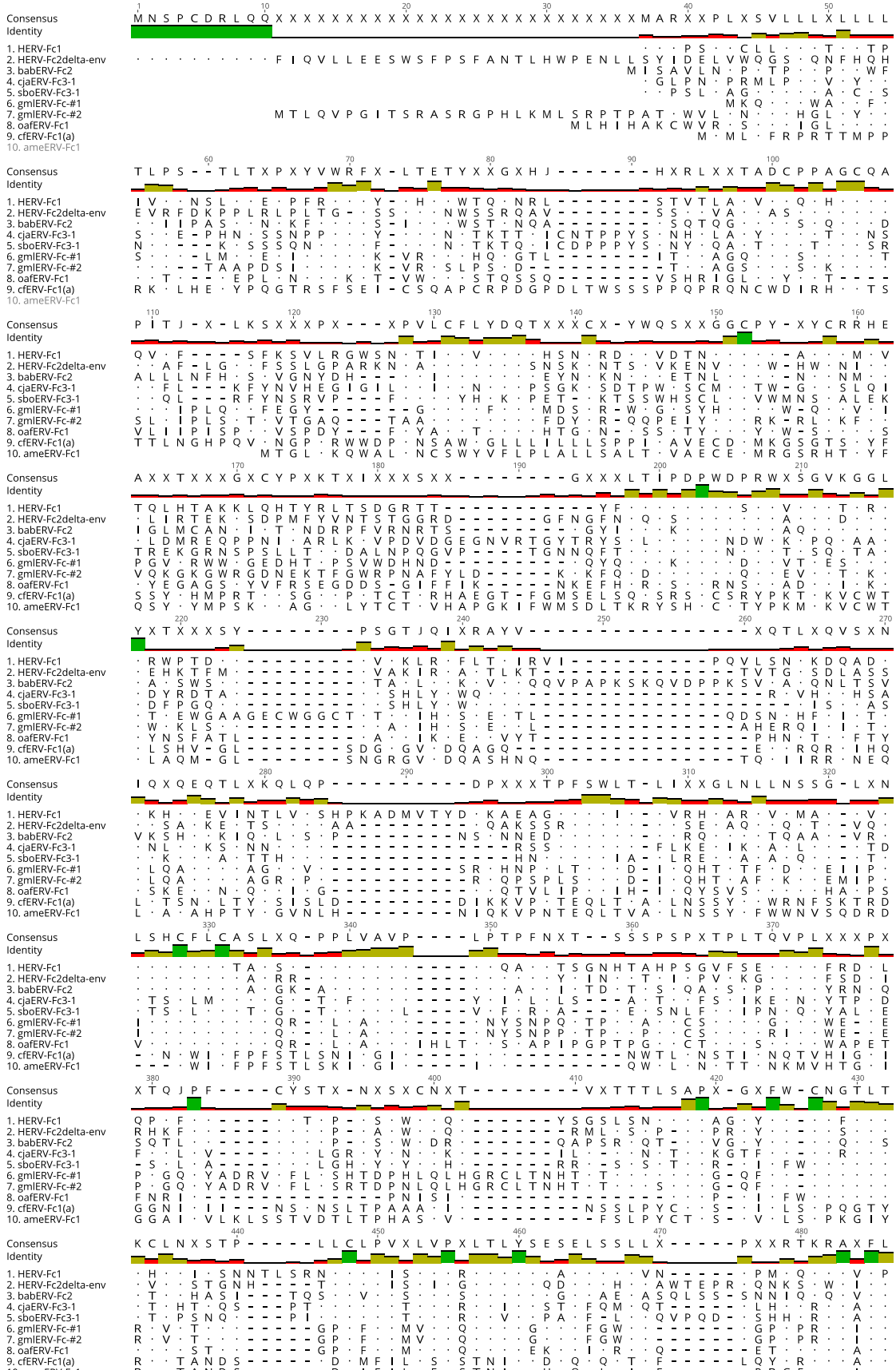
Figure 3.1: Schematic representation of ERV-Fc proviruses with *env* ORFs. Position of *gag, pol, env* and LTRs within the provirus is indicated above or below the element, ORF *envs* are shaded medium gray (classical ERV-Fc *env*) or light green (recombinant *env*). Green line=frame shift, red line=early stop, ??=unknown, NNN=N's in sequence, Purple line=insertion, Δ=deletion. On the left species are listed that the provirus is found in and its specific name. The right lists the accession number, chromosome or contig position and the LTR identity determined through pairwise alignments in Geneious.

Species	Species	ERV Name		LTR % Identity	Accession #	Position
Human	*Homo sapien*	HERV-Fc1-env		92.8	NC_000023.11	97841482-97849424
Human	*Homo sapien*	HERV-Fc2Δ-env		N/A	NC_000007.14	153414153-153409357
Chimpanzee	*Pan troglodytes*	homERV-Fc1-env		91.8	NC_006491.4	97450004-97457999
Bonobo	*Pan paniscus*	ppaERV-Fc1-env		91.8	NC_027891.1	97220842-97228837
Baboon	*Papio anubis*	babERV-Fc2-env		100	AC091754.4	87145-95671
Marmoset	*Callithrix jacchus*	cjaERV-Fc3-1-env		90.3	NC_013901.1	16020741-16025666
Squirrel Monkey	*Saimiri boliviensis boliviensis*	sboERV-Fc3-1-env		N/A	NW_003943680.1	10414173-10407491
Grey Mouse Lemur #1	*Microcebus murinus*	gmlERV-Fc-#1-env		99.3	NC_033679.1	13434948-13443777
Grey Mouse Lemur #2	*Microcebus murinus*	gmlERV-Fc-#2-env		83.7	NC_033670.1	92528996-92523310
Aardvark	*Orycteropus afer afer*	oafERV-Fc1-env		N/A	ALYB01310164.1	1-~3650
Dog	*Canis lupus familiaris*	CfERV-Fc1(a)-env		98.5	NW_003726126.1	50566103-50573966
Panda	*Ailuropoda melanoleuca*	ameERV-Fc1-env		89.1	NW_003218401.1	204452-196924

ambiguities prevented determining whether the cjaERV-Fc3-1, oafERV-Fc1 and ameERV-Fc1 Env ORFs are found in full-length proviruses (Fig. 3.1). In sboERV-Fc3-1 there is a large deletion of the 5' end of the provirus including the 5'-LTR and part of *gag.* I could not determine whether sboERV-Fc3-1 has an ORF in *pol* due to sequence ambiguities in the squirrel monkey genome assembly. There is a large deletion between *gag* and *pol* in gmlERV-Fc-#2 (Fig. 3.1). Besides the above mentioned proviral defects, all ERV-Fc *gag* and *pol* genes had either premature stop codons (red lines), frameshift mutations (green lines) or deletions (orange lines). The 5'-3' LTR identities ranged from 83.7% (gmlERV-Fc-#2) to 100% (babERV-Fc2) (Fig. 3.1).

The ERV-Fc Envs from different species have low identity in the SU domain

An alignment was made to gain further insight into ERV-Fc Envs (Fig. 3.2). The predicted ERV-Fc Env proteins have motifs and features typical of exogenous gammaretroviruses (Fig. 3.2 and 3.3A). They code for a surface (SU) subunit and a transmembrane (TM) subunit. Within SU, a signal peptide (SP) is predicted for all the sequences except HERV-FcΔenv; there are between 5 and 12 predicted glycosylation sites and a conserved CXXC isomerization domain (Fig. 3.3A). In TM, there is a predicted fusion peptide immediately after the furin cleavage site, a conserved immunosuppressive domain (ISD), a CX_6CC domain, and a membrane spanning domain (MSD) (Fig. 3.3A). The predicted amino acid sequences have a range of identity between 18.8% and 70.5% and a similarity between 35.15% and 76.53% (Fig. 3.3A and B). cfERV-Fc1(a) and ameERV-fc1 have an identity of 68% and similarity of 78.53%, but have low similarity and


Consensus
Identity
1. HERV-Fc1
2. HERV-Fc2delta-env
3. babERV-Fc2
4. cjaERV-Fc3-1
5. sboERV-Fc3-1
6. gmlERV-Fc-#1
7. gmlERV-Fc-#2
8. oafERV-Fc1
9. cfERV-Fc1(a)
10. ameERV-Fc1

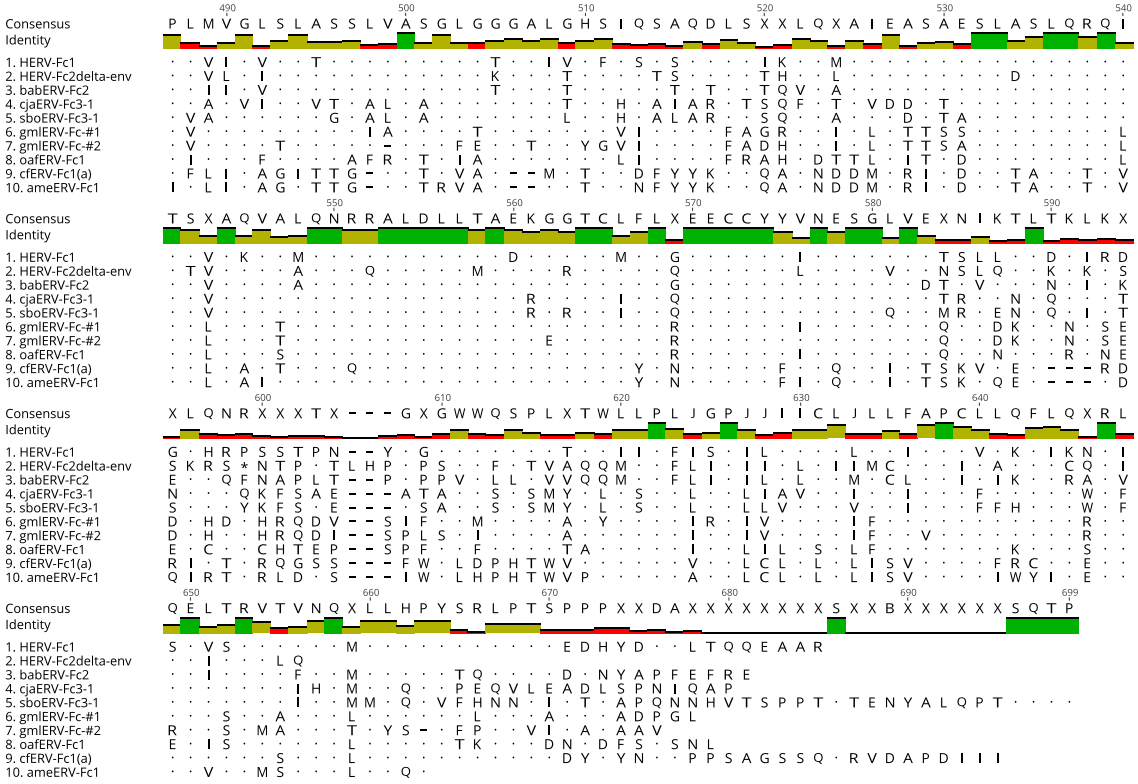
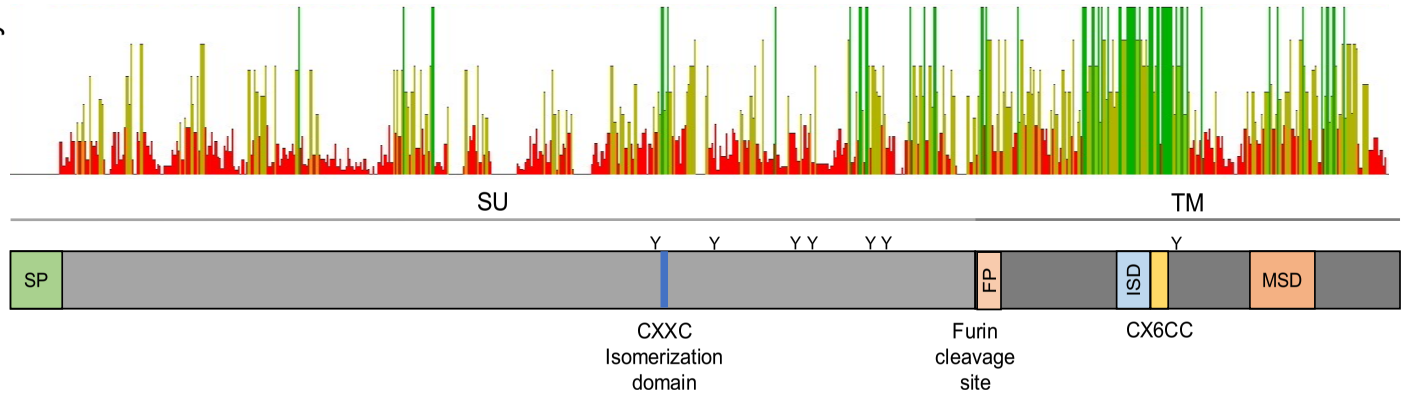


Figure 3.2: Alignment of the ERV-Fc Envs. Alignment made in Geneious used to generate data for Fig. 3.1. Disagreements to the consensus are highlighted.

Figure 3.3: Conservation of the ERV-Fc Envs is low in the SU domain and higher in the TM domain. (A) Conserved features of the ERV-Fc Envs, including the SU (light gray) and TM (dark gray) domains. Within the SU is the signal peptide (SP), the CXXC isomerization domain, and six predicted N-glycan sites, indicated by a Y, that are conserved in at least five of the sequences. Separating the TM and SU is the furin cleavage site. In the TM positions of the fusion peptide (FP), immunosuppressive domain (ISD), CX6CC motif, and the membrane spanning domain (MSD) are indicated. There is also one N-glycan site that is conserved in the Envs. Above is a graph representing the average pairwise percent amino acid identity of the Envs. Green indicates 100% identity, yellow 30%>100% identity, red shows >30% identity. Height of the graph at each position indicates the fraction of amino acids that are identical at that position. Blanks illustrate where there is only one sequence in the alignment at that position. (B) Tables with the pairwise percent amino acid similarity (Blosum62 with threshold 1) of the full length Envs, the SU domain of the Envs and the TM domain of the Envs. Lighter green indicates less similarity and darker green indicates higher similarity.

A



B

ERV-Fc-Env Full length % Similarity

Env	HERV-Fc1	HERV-Fc2Δenv	babERV-Fc2	cjaERV-Fc3-1	sboERV-Fc3-1	gmlERV-Fc-#1	gmlERV-Fc-#2	oafERV-Fc1	cfERV-Fc1(a)	ameERV-Fc1
HERV-Fc1		58.8	62.21	51.68	49.75	54.62	50	50.34	37.95	38.48
HERV-Fc2Δenv	58.8		58.6	53.52	52.3	49.12	47.22	49.82	35.15	36.69
babERV-Fc2	62.21	58.6		51.33	51.75	53.69	51.34	51.01	37.69	37.67
cjaERV-Fc3-1	51.68	53.52	51.33		70.91	48.4	46.77	47.52	36.14	37.93
sboERV-Fc3-1	49.75	52.3	51.75	70.91		49.16	46.93	47.68	34.95	37.3
gmlERV-Fc-#1	54.62	49.12	53.69	48.4	49.16		76.4	58.97	36.41	36.81
gmlERV-Fc-#2	50	47.22	51.34	46.77	46.93	76.4		58.62	35.2	35.4
oafERV-Fc1	50.34	49.82	51.01	47.52	47.68	58.97	58.62		38.25	39.39
cfERV-Fc1(a)	37.95	35.15	37.69	36.14	34.95	36.41	35.2	38.25		78.53
ameERV-Fc1	38.48	36.69	37.67	37.93	37.3	36.81	35.4	39.39	78.53	

% Similarity
0-10
10-20
20-30
30-40
40-50
50.1-60
60-70
70-80
80-90
90-99
100

ERV-Fc-Env SU % Similarity

Env	HERV-Fc1	HERV-Fc2Δenv	babERV-Fc2	cjaERV-Fc3-1	sboERV-Fc3-1	gmlERV-Fc-#1	gmlERV-Fc-#2	oafERV-Fc1	cfERV-Fc1(a)	ameERV-Fc1
HERV-Fc1		52.28	53.12	44	42.79	44.39	41.04	38.29	24.94	24.92
HERV-Fc2_env	52.28		48.48	46.7	44.39	39.9	38.19	41.84	24.81	24.84
babERV-Fc2	53.12	48.48		43.6	45.02	45.02	44.8	40.85	25.98	23.81
cjaERV-Fc3-1	44	46.7	43.6		64.12	40.15	38.89	37.95	24.61	24.13
sboERV-Fc3-1	42.79	44.39	45.02	64.12		40.7	38.58	39.38	24.48	24.21
gmlERV-Fc-#1	44.39	39.9	45.02	40.15	40.7		70.79	47.93	23.13	22.49
gmlERV-Fc-#2	41.04	38.19	44.8	38.89	38.58	70.79		50	24.24	22.74
oafERV-Fc1	38.29	41.84	40.85	37.95	39.38	47.93	50		24.48	25.32
cfERV-Fc1(a)	24.94	24.81	25.98	24.61	24.48	23.13	24.24	24.48		71.29
ameERV-Fc1	24.92	24.84	23.81	24.13	24.21	22.49	22.74	25.32	71.29	

ERV-Fc-Env TM % Similarity

Env	HERV-Fc1	HERV-Fc2Δenv	babERV-Fc2	cjaERV-Fc3-1	sboERV-Fc3-1	gmlERV-Fc-#1	gmlERV-Fc-#2	oafERV-Fc1	cfERV-Fc1(a)	ameERV-Fc1
HERV-Fc1		73.56	80.71	67.35	63.68	75.77	68.75	74.87	64.18	63.69
HERV-Fc2Δenv	73.56		81.61	68.97	70.11	70.11	68.97	67.82	58.62	58.62
babERV-Fc2	80.71	81.61		67.35	65.48	71.65	65.1	71.79	61.93	63.69
cjaERV-Fc3-1	67.35	68.97	67.35		84.1	65.46	63.02	66.67	58.97	62.36
sboERV-Fc3-1	63.68	70.11	65.48	84.1		66.49	64.06	64.1	54.29	60.67
gmlERV-Fc-#1	75.77	70.11	71.65	65.46	66.49		87.5	80.93	63.92	63.13
gmlERV-Fc-#2	68.75	68.97	65.1	63.02	64.06	87.5		76.04	57.81	58.1
oafERV-Fc1	74.87	67.82	71.79	66.67	64.1	80.93	76.04		65.64	64.25
cfERV-Fc1(a)	64.18	58.62	61.93	58.97	54.29	63.92	57.81	65.64		91.28
ameERV-Fc1	63.69	58.62	63.69	62.36	60.67	63.13	58.1	64.25	91.28	

identity in comparison to the rest of the ERV-Fc Envs due to a recombination event that replaced a portion of SU with a heterologous SU sequence (Fig. 3.1 and Fig. 3.3A and B) [77]. Across retrovirus families, the SU subunit is typically less conserved than is the TM subunit. This is consistent with our findings, for which the range is 8.4% to 64.7% (identity) and 22.49% to 71.29% (similarity) in SU, and 35.6% to 82.6% (identity) and 54.29% to 91.28% (similarity) in TM (Fig. 3.2 and 3.3A and B).

Intact ERV-Fc Envs are defective for fusion and entry

We next assessed expression and functionality of ten ERV-Fc Env proteins by western blot (Fig. 3.4A). Protein was detected for all constructs in varying amounts; three had consistently very low levels of protein expression (HERV-FcΔenv, gmlERV-Fc-#1 and ameERV-Fc1), and conERV-Fc1 (a consensus between the human, chimpanzee, bonobo and gorilla sequences) was expressed at the highest levels (Fig. 3.4A and B). Processing by furin cleavage between the SU and TM was not observed for most of the Envs. While this was expected for HERV-Fc2Δenv, babERV-Fc2 and ameERV-Fc1, which do not have intact furin cleavage sites, it was unexpected for several of the others, which retain intact furin target motifs (Fig. 3.4A and C). Even with intact furin target sites, no processing was observed for cjaERV-Fc3-1 and sboERV-Fc3-1, and little to none for conERV-Fc1. Only gmlERV-Fc-#2, oafERV-Fc1 and cfERV-Fc1(a) were processed between SU and TM with any efficiency (Fig. 3.4A and C). All of these cleavage defects indicate that furin cleavage has not been

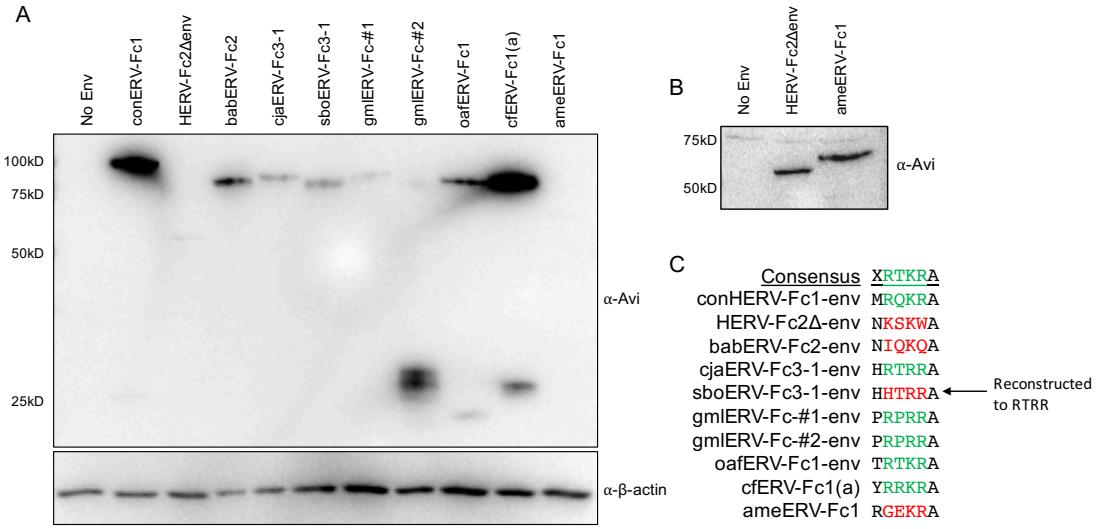


Figure 3.4: All ERV-Fc Envs are expressed but not fully processed. (A) Western blot analyses of cell lysate from 293T/17 cells transfected with expression plasmids containing the codon optimized Avi tagged ERV-Fc *env* ORFs. (B) Re-run western blot with more protein for low expressing ERV-Fc Envs HERV-Fc2Δenv and ameERV-Fc1. (C) Alignment of ERV-Fc Envs furin cleavage sites. Green=canonical/intact cleavage site, Red=mutated cleavage site.

preserved in the ERV-Fc Envs. Furin cleavage defects would also inhibit ability to drive membrane fusion and viral entry into a host cell.

Restoration of a signal peptide to HERV-Fc2Δenv leads to glycosylation

HERV-FcΔenv is expressed, but gives a band smaller than the predicted size (Fig. 3.4 A and B). With 13 possible N-linked glycosylation sites, the expected observed size is ~85.5kD; the observed size of ~60.4kD is close to the predicted size in the absence of glycosylation (Fig.3A and B). A possible explanation for the discrepancy is that HERV-FcΔenv is not trafficking through the secretory pathway, which prevents the Env from being glycosylated. To investigate this possibility, we first checked for the presence of a signal peptide in the predicted HERV-FcΔenv ORF. Examination of the conERV-Fc1, babERV-Fc2 and HERV-FcΔenv sequence by the SignalIP 4.1 server indicates that conERV-Fc1 has a predicted signal peptide and cleavage site between positions 22 and 23, and babERV-Fc2 has a signal peptide and predicted cleavage site between position 28 and 29. HERV-FcΔenv, however, is not predicted to have a signal peptide (Fig. 3.5 A, B and C) [137]. To test this possibility, we added two different heterologous signal peptides to the HERV-Fc2Δenv construct, based on the predicted SPs of babERV-Fc2 and conERV-Fc. When the chimeric constructs were tested by transfection and western blot, the observed proteins ran at ~85.5kD, consistent with glycosylation and trafficking through the secretory pathway (Fig. 3.5 D).

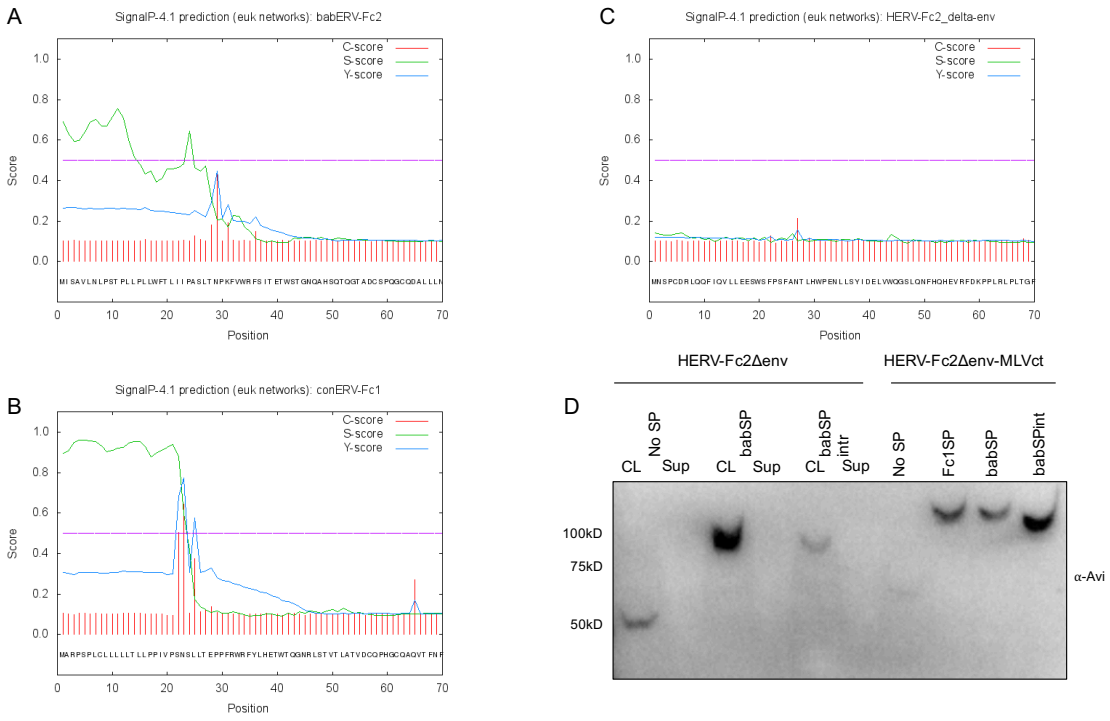


Figure 3.5: HERV-Fc2Δenv does not have a signal peptide, adding the conERV-Fc1 or babERV-fc2 signal peptide to HERV-Fc2Δenv leads to glycosylation. (A) babERV-Fc2 predicted signal peptide (B) conERV-Fc1 predicted signal peptide (C) HERV-Fc2Δenv predicted signal peptide. C-score (red) is a prediction of where the signal peptide cleavage site is. S-score (green) is the signal peptide score, predicts the location of the entire signal peptide. Y-score (blue) combined cleavage site score, combines the C and S score to more accurately predict where the signal peptide cleavage site is placed. The higher the score the higher the probability of a signal peptide and cleavage site. Data generated using SignalIP 4.1 server [137]. (D) Western blot analyses of cell lysate (CL) and supernatant (Sup) from 293T/17 cells transfected with expression plasmids containing the HERV-Fc2Δenv signal peptide chimeric constructs.

Reconstruction of the furin cleavage site does not rescue SU-TM cleavage of HERV-Fc2Δenv

Reconstruction of the HERV-Fc2Δenv furin cleavage site had no effect on processing of TM and SU (Fig. 3.6B). Examination of the predicted amino-acid sequence revealed a putative N-linked glycosylation motif that overlaps the predicted furin-cleavage site and that is not found in other ERV-Fc Env, suggesting post-translation glycosylation could be interfering with processing (Fig. 3.6A). To test this possibility, a construct with the babERV-Fc2 *env* SP and a reconstructed furin cleavage site (modified to remove the N-glycosylation site) was made (Fig. 3.6A). However, elimination of the N-linked glycosylation site at the furin cleavage site (NKSKR to NRQKR) did not result in detectable cleavage of the precursor protein, indicating there may be other changes outside of the motif that affect recognition or the cleavage by furin protease (Fig. 3.6B).

Distal mutants fail to restore furin cleavage of conERV-Fc1

Even though conERV-Fc1 has a canonical furin cleavage site, minimal processing is observed, similar to reports of reconstructed Refrex-1 (Fig. 3A) [131]. Multiple mutational strategies were attempted to improve cleavage. When the ERV-Fc Envs furin cleavage sites were aligned along with the two reconstructed Refrex-1s, several major differences were found (Fig. 3.7A) [131]. At residue 387 in conERV-Fc, immediately downstream of the cleavage site, there is a proline (P) in place of a highly-conserved leucine (L) (Fig. 3.7A). Immediately upstream of the cleavage site, babERV-Fc2 Env has 5 extra amino acids not present in conERV-Fc1 Env (site 375). In the area upstream of the

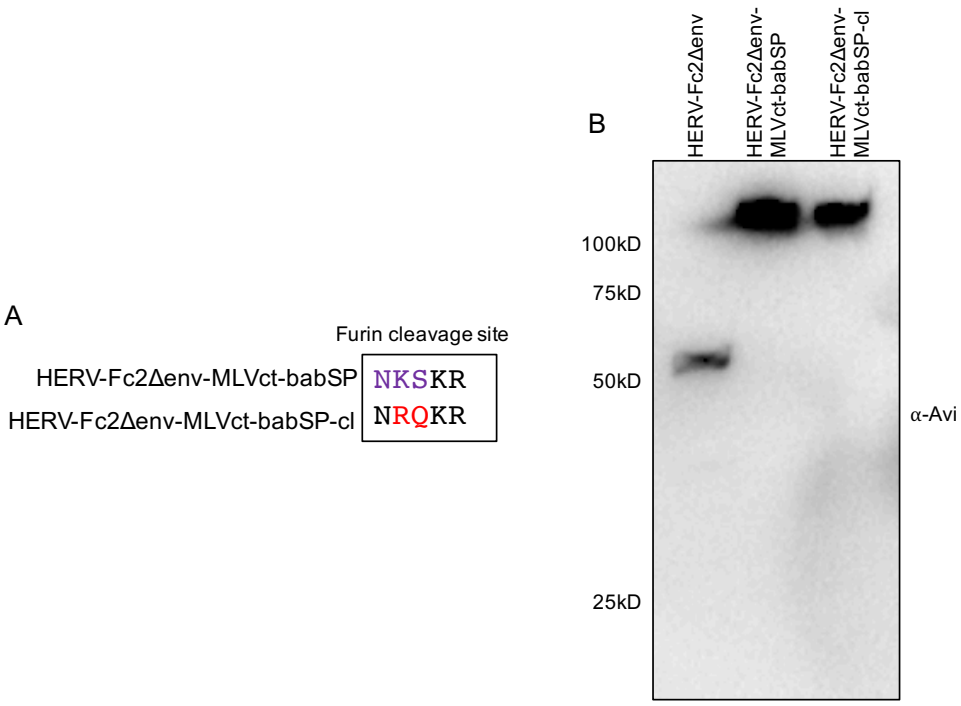


Figure 3.6: Removing the N-glycosylation site in HERV-Fc2Δenv does not restore furin cleavage. (A) Alignment of HERV-Fc2Δenvs furin cleavage site, purple = N-glycosylation site, red = residues changed to remove glycosylation site. (B) Western blot analysis of cell lysate harvested from 293T/17 cells transfected with Avi-tagged HERV-Fc2Δenv constructs.

A

```
Consensus APXGXXFWCNGTLTKCLNXSQ----TXGLCLPVXLVPXLTLYSESEFSXLLX-----PXXRTRRAXF-LP
ERV-DC7   ...R..WV..TG..P.IS-L.VLNI..DY.ILI..W..I-F.HD..YI.GH.EPGS----.F..E.VS.T
ERV-DC16  ......WV..TG..P.IS-L.VLNT..DD.ILI..W..I-F.HD..YI.GH.EPGS----.F..E.VS.T
conHERV-Fc1-env ..........F....H....SNNTLS.N....I..........A.L...V.-----....QK.....-P.
HERV-Fc2Δ-env ...R......S....V...TGNH----T....I..I.......QD.L.....WTEPR...KSKW....-..
babERV-Fc2-env ..........Q..S.T...ASI----..S..V.............G.LA..A.QLSS-S..IQKQ....-..
cjaERV-Fc3-1-env ..........R....T.H...S---P.L.............I..ST..Q..Q.-----.....-------..
sboERV-Fc3-1-env ...............T.P.N.P---P.L.............V..PA..L..Q.PQD----...H.......-..
gmlERV-Fc-#1-env ..-............R.V..--------...P.F...V.......G....G..----------..P.....-..
gmlERV-Fc-#2-env ..-............R.V..--------...P.F...V.......G....G..----------..P.....-..
oafERV-Fc1-env ...................T-------...P.F...........EK..I.----------...K.....-..
```

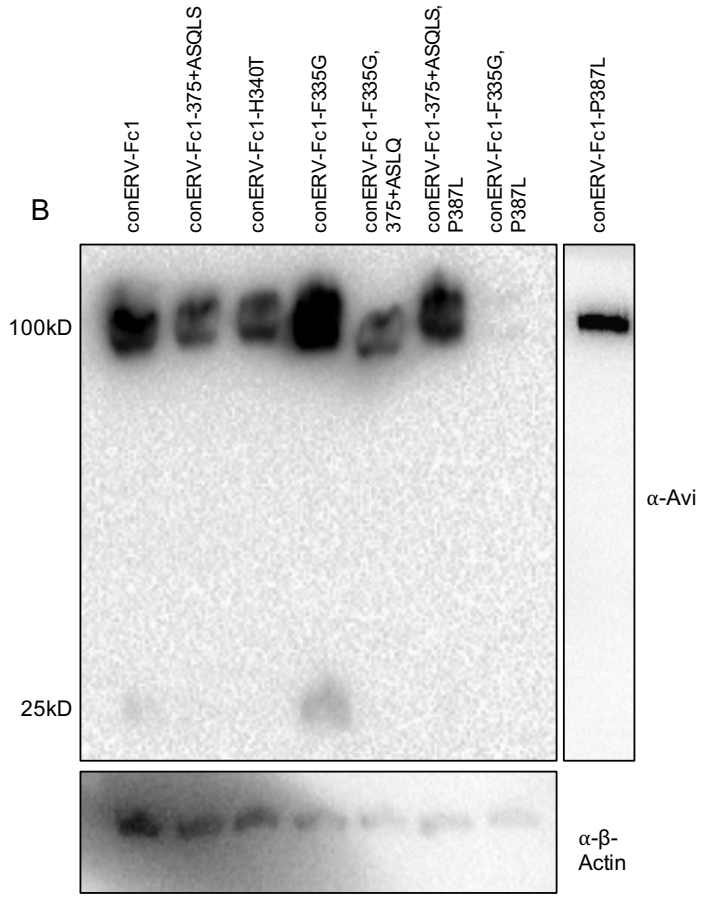


Figure 3.7: Cleavage mutants do not restore furin cleavage to conERV-Fc1. (A) Alignment of non-recombinant ERV-Fc, ERV-DC7 and ERV-DC16 Envs furin cleavage site and surrounding sequence. Furin cleavage site in green, changes made in ERV-DC7 and ERV-DC16 to restore furin cleavage (Ito *et al.* 2016). Residues changed in conERV-Fc1 to attempt to restore furin cleavage are highlighted in red. (B) Western blot analysis of lysate of 293T/17 cells transfected with conERV-Fc1 furin cleavage mutants, all constructs are Avi tagged.

furin cleavage site, shown to be important for Refrex-1 cleavage [131], there are two sites of interest in conERV-Fc1. At site 335 in conERV-Fc1, there is a hydrophobic phenylalanine (F), which in the other Envs is a neutral site, most commonly a glycine (G) (Fig. 3.7A). Finally, at site 340 there is histidine (H) in conERV-Fc1 Env, again very different from the conserved cysteine (C) or threonine (T) found at that location in the ERV-Fc Envs (Fig. 3.7A). We therefor replaced these elements in various combinations and tested the modified constructs for expression and processing by transfection and western blot (Fig. 3.7A and B). However, no significant improvement in furin cleavage was observed compared to the parental construct (Fig. 3.7B).

Reconstruction of babERV-Fc2 Env restores processing and infection capability

The babERV-Fc2 is the youngest of the ERV-Fc proviruses with an intact *env* ORF, estimated at ≤1 million years old [74]. Env expression is detectable by transient transfection, but the resulting protein is not cleaved into separate SU and TM domains (Fig. 3.4A). The native sequence lacks a canonical furin cleavage site, most likely due to mutations acquired subsequent to endogenization (Fig. 3.4C and 3.8). To test this possibility, the furin cleavage site was first repaired through site directed mutagenesis (IQKQ to RQKR) to produce babERV-Fc2-cl (where cl indicates the modification to contain a canonical furin-cleavage motif) (Fig. 3.8). This change in babERV-Fc2-cl resulted in proper cleavage of the precursor molecule, as judged by the

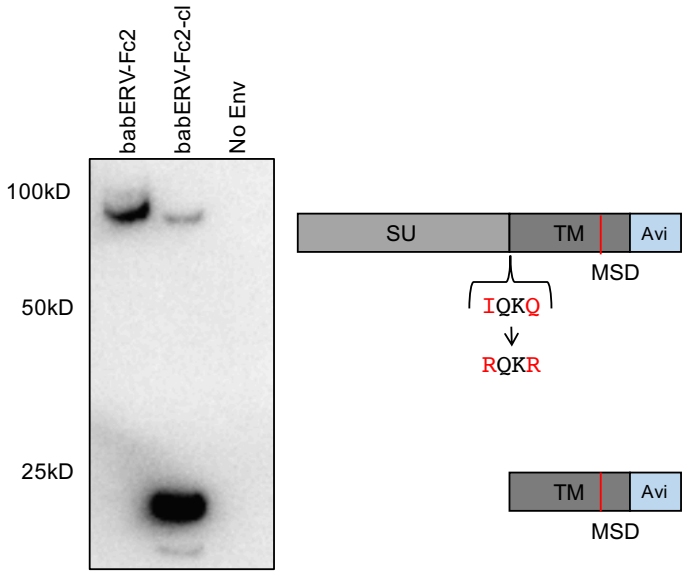


Figure 3.8: Reconstructed babERV-Fc2 furin cleavage site is functional. Western blot analysis of babERV-Fc2 and babERV-Fc2-cl (α-Avi). To the right is a schematic of what portion of the Env is seen on the blot: top-full length Env, bottom-TM subunit. Under the full length Env is the babERV-Fc2 furin cleavage site, with an arrow pointing at the reconstructed cleavage site in babERV-Fc2-cl, red amino acids are the ones that were changed.

appearance of a band corresponding to the Avi-tagged TM subunit on western blots (Fig. 3.8).

I next asked whether the modified babERV-Fc2-cl protein was functional. To determine functionality, MLV viral particles were pseudotyped with babERV-Fc2-cl and used to infect 293T/17 cells. Little to no infection was seen (Fig. 3.10), indicating that the Env was not functional, or that the cells do not express the receptor. However, the Env proteins of many gammaretroviruses are known to have a C-terminal R-peptide [18], [79], [82]. The presence of the R-peptide prevents fusion until it is cleaved off by the viral protease during the maturation stage of viral replication (thereby activating the Env protein's fusogenic ability) [82]. The R-peptide cleavage site of MLV is within an L-V motif in the cytoplasmic tail – a motif that is not present in babERV-Fc2-cls cytoplasmic tail (Fig. 3.9A). We therefore hypothesized that babERV-Fc2-cl has an R-peptide, but either the MLV protease does not recognize the ERV-Fc R-peptide cleavage site or the babERV-Fc2 R-peptide domain has lost the cleavage site. To determine whether babERV-Fc2 Env has an R-peptide, we made four truncations of the babERV-Fc2-cl cytoplasmic tail: babERV-Fc2-cl-Δ4AA, babERV-Fc2-cl-Δ12AA, babERV-Fc2-cl-Δ22AA and babERV-Fc2-cl-Δ29AA (Fig. 3.9A). A chimeric Env was also constructed where the babERV-Fc2-cl cytoplasmic tail was replaced with the MLV *env* cytoplasmic tail (babERV-Fc2-cl-MLVct), which contains an R-peptide cleavage site that is recognized and cleaved by the MLV protease (Fig. 3.9A).

A

MLV-ct	LVQFVKDRISVVQALVLTQQ----YHQLKPLEYEP
babFc2-MLVct	LIKFLKDRISVVQALVLTQQ----YHQLKPLEYEP-Avi
babFc2-ct	LIKFLRARVQEITRVTFNQMLLHPYTQLPTSDPNYAP-Avi
Δ4AA	LIKFLRARVQEITRVTFNQMLLHPYTQLPTSDP-Avi
Δ12AA	LIKFLRARVQEITRVTFNQMLLHPY-Avi
Δ22AA	LIKFLRARVQEITRV-Avi
Δ29AA	LIKFLRAR-Avi

B

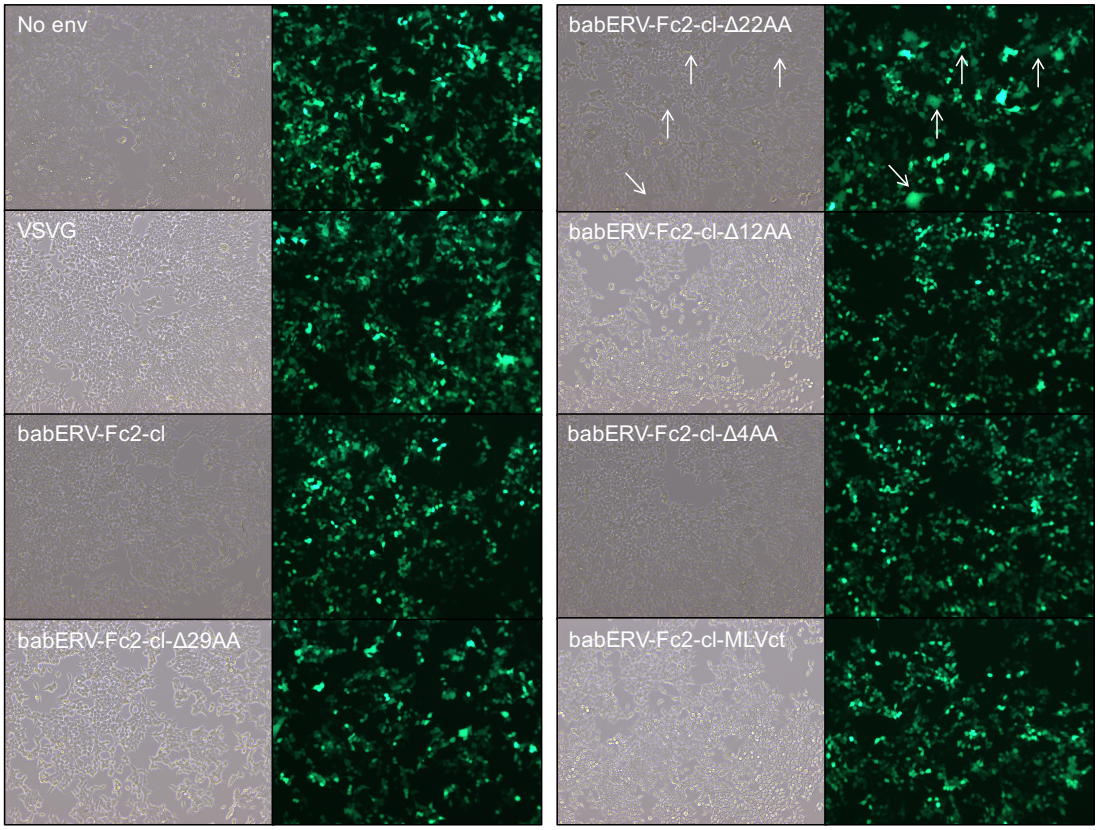


Figure 3.9: Truncation of 22AA off the babERV-Fc2 cytoplasmic tail leads to syncytia formation in 293T/17 cells. (A) Alignment of the MLV cytoplasmic tail (ct) (orange), the babERV-Fc2 ct (black), the chimeric babERV-Fc2 Env and the four truncation mutants. The first arrow indicates the MLV R-peptide cleavage site. The Second arrow indicates the tyrosine motif. (B) Cell-cell fusion in 293T/17 cells transfected with babERV-Fc2 constructs and with VSVG, arrows indicate syncytia. Cells co-transfected with GFP expression vector.

Ectopic expression of babERV-Fc2-cl-Δ22AA in 293T/17 cells led to observable syncytia formation (Fig. 3.9B), whereas babERV-Fc2-cl-Δ4AA, babERV-Fc2-cl-Δ12AA and babERV-Fc2-cl-Δ29AA did not (Fig. 3.9B). This suggests that deletion of the C-terminal 22 residues removed a putative R-peptide and activated the inherent fusogenicity of the babERV-Fc2 Env protein. Surprisingly, the babERV-Fc2-cl-Δ29AA construct, which would also eliminate the R-peptide, did not have the same effect. This may be due to several unknown factors. For example, the conformation of the Env may have been altered, affecting the Envs ability to cause cell-cell fusion. Alternatively, the cytoplasmic tail is important for proper trafficking within a host cell [138], and therefore babERV-Fc2-cl-Δ29AA may not have trafficked properly, which could have led to a reduction in the amount of Env at the cell surface and reduced cell-cell fusion.

MLV particles pseudotyped with babERV-Fc2-cl-Δ22AA were able to infect 293T/17 cells (Fig. 3.10). This indicates that 293T/17 cells express the appropriate receptor, and that babERV-Fc2-cl-Δ22AA is able to mediate fusion between membranes. MLV particles pseudotyped with babERV-Fc2-cl-MLVct were also able to infect 293T/17 cells at a higher rate than virus pseudotyped with babERV-Fc2-cl-Δ22AA, indicating the MLV protease can cleave off the R-peptide and confirming that there is a possible receptor on 293T/17 cells (Fig. 3.9). The decreased rate of infection with babERV-Fc2-cl-Δ22AA may be due to a reduced incorporation rate into virions. It has been shown that a tyrosine (Y) motif in the cytoplasmic tail is important for incorporation of the Env into virions [139]. This motif was removed in babERV-Fc2-cl-Δ22AA (Fig. 3.9A).

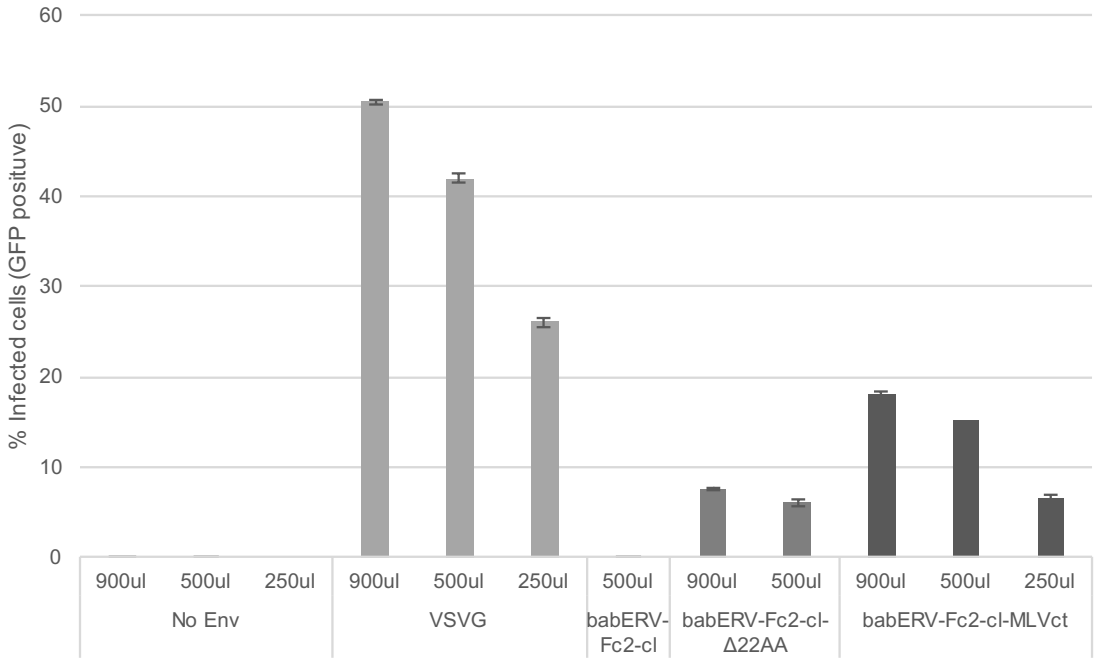


Figure 3.10: A repaired and truncated babERV-Fc2 can mediate viral infection. Infection of 293T/17 cells with MLV particles pseudotyped with babERV-Fc-cl, babERV-Fc2-Δ22AA, babERV-Fc-cl-MLVct or controls VSVG and No Env. N=3.

Alternatively, the formation of syncytia in cells expressing babERV-Fc2-cl-Δ22AA may have caused cell death resulting in a decreased number of viral particles forming.

To gain insight into the cell tropism of the potential babERV-Fc2, multiple cell lines across several species were tested for infectivity. MLV particles pseudotyped with babERV-Fc2-cl-MLVct (MLV-babERV-Fc2-cl-MLVct) were used to infect 293T/17 cells (human), HT1080 cells (human), HOS cells (human), A549 cells (human), Vero cells (African green monkey (AGM)), LLC-MK2 cells (rhesus), MDCK cells (canine), CRFK cells (feline) and DF-1 cells (chicken) (Fig. 3.11 and 3.12). MLV-babERV-Fc2-cl-MLVct was able to infect all the human cell lines tested, but infection was not detectable in the other, nonhuman cell lines (Fig. 3.11 and 3.12), suggesting the presence of a specific receptor on human cells but not on the others tested. This could mean that non-human cells lack the receptor or, alternatively, babERV-Fc2-cl-MLVct does not bind to non-human orthologs of the receptor.

Both gmlERV-Fc-#2 and oafERV-Fc1 are processed into SU and TM (Fig. 3.4A), indicating they could retain canonical viral functions. Chimeric Envs containing the MLV-ct were made to test infectivity of gmlERV-Fc-#2-MLVct and oafERV-Fc1-MLVct against a similar panel of cells as babERV-Fc2-cl-MLVct; however, very little to no infection was observed (Fig. 3.12). Concentrating pseudotyped virus also did not result in detectable infection (Fig. 3.13) These results suggest that either the Env proteins are not functional, or that none of the cells tested express an appropriate cell-surface receptor.

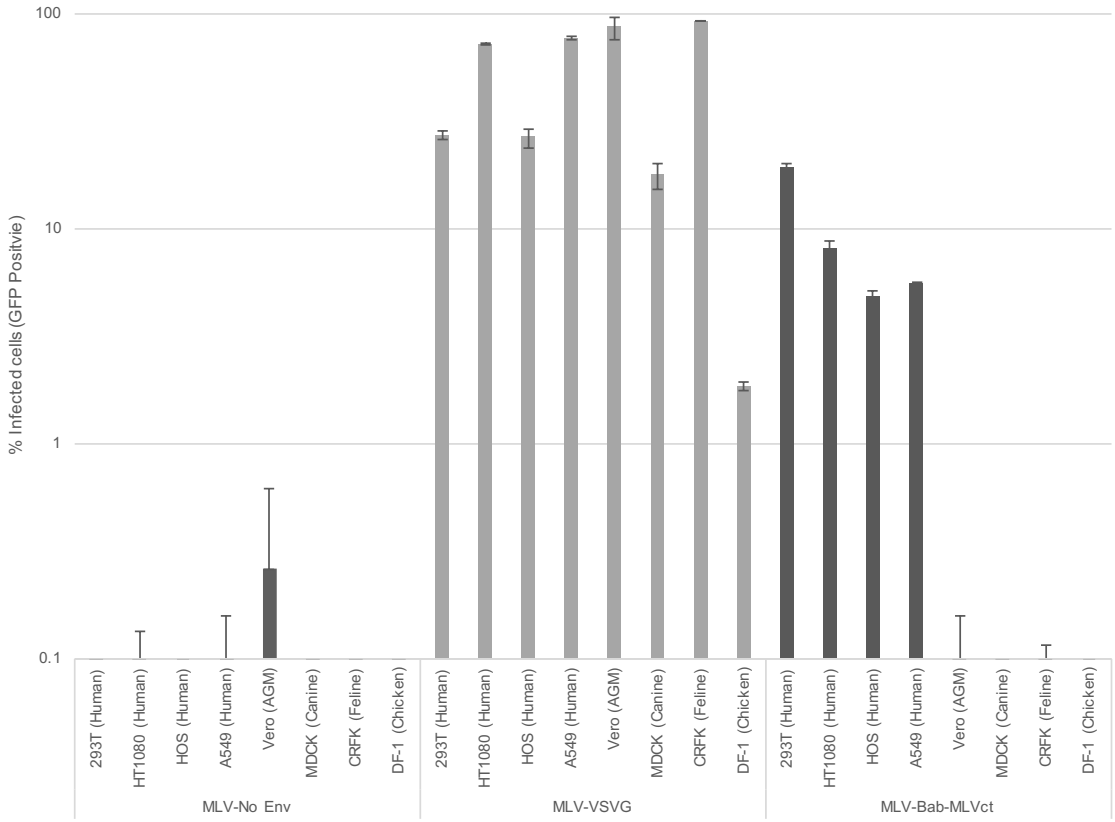


Figure 3.11: Human cell lines have a receptor used by babERV-Fc2-cl-MLVct. Infectivity of MLV particles pseudotyped with the babERV-Fc2 chimeric Env or controls VSVG and No Env containing a GFP reporter, log scale, N=3.

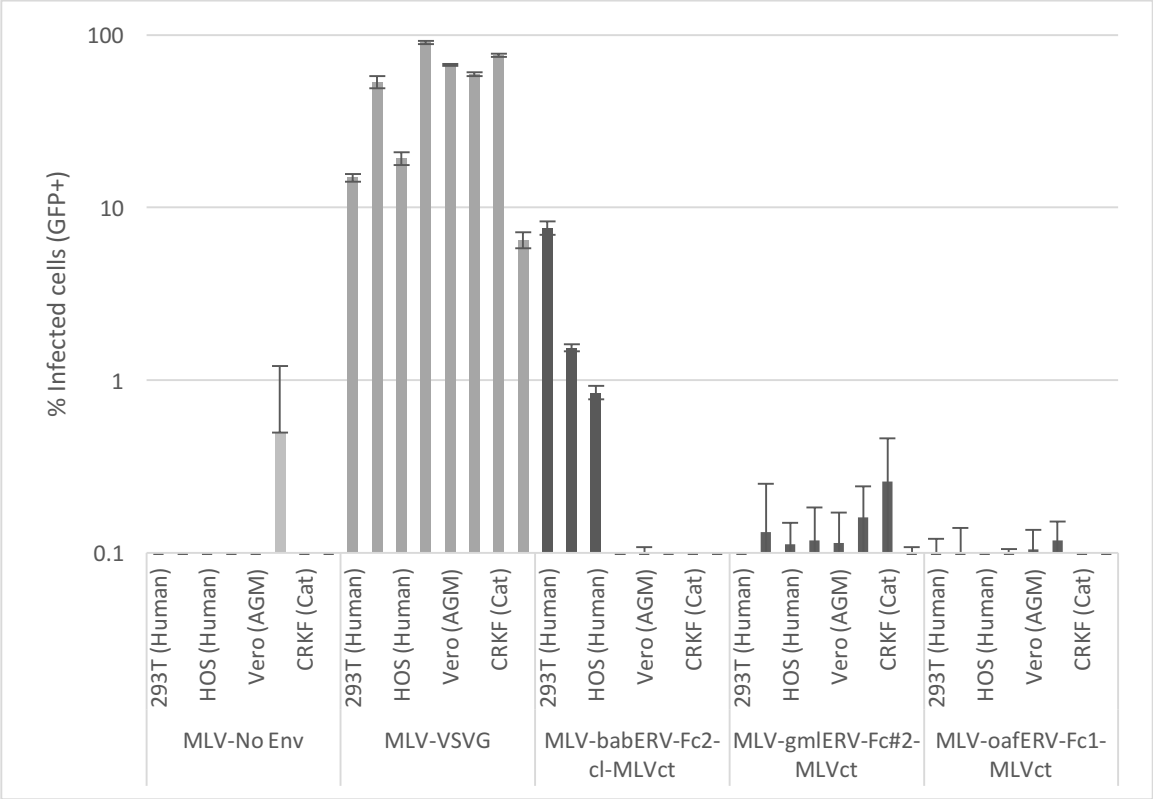


Figure 3.12: Human cell lines have a receptor used by babERV-Fc2-cl-MLVct but not gmlERV-Fc#2-MLVct and oafERV-Fc1-MLVct. Infectivity of MLV particles pseudotyped with the babERV-Fc2, gmlERV-Fc#2 or oafERV-Fc1 chimeric Envs or controls VSVG and No Env containing a GFP reporter on a panel of cells. log scale, N=3.

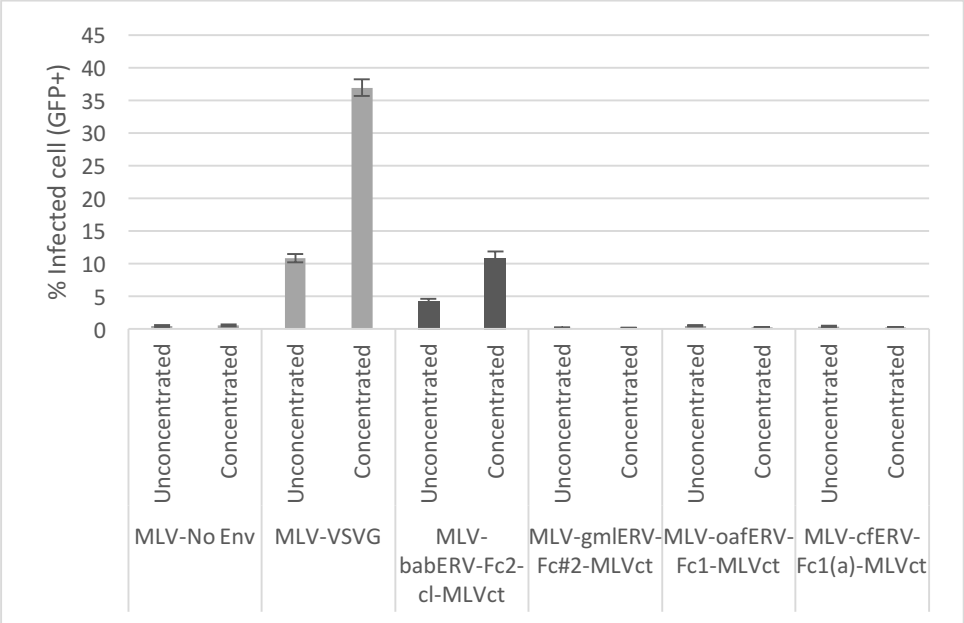


Figure 3.13: Lemur, aardvark and dog ERV-Fc Envs do not infect 293T/17 cells. Infection of 293T/17 cells with MLV pseudotyped with lemur, aardvark or dog ERV-Fc Envs with MLVcts, controls No Env, VSVG and babERV-Fc2-cl-MLVct. Virus was left unconcentrated or was concentrated before infections. N=2.

babERV-Fc2 Env confers superinfection resistance to virus pseudotyped with babERV-Fc2-cl-MLVct

To test the antiviral potential of babERV-Fc2 and the other ERV-Fc Envs, a superinfection interference assay was performed. As negative controls, 293T/17 cells were transfected with empty vector or a vector expressing SIVgp160. SIV uses CD4 as a primary receptor, which is not expressed on 293T/17 cells. In parallel, cells were transfected with the ERV-Fc *envs*: conERV-Fc1, HERV-Fc2Δenv, HERV-FcΔenv-MLVct-babSP, babERV-Fc2, babERV-Fc2-cl-MLVct, cjaERV-Fc3-1, sboERV-Fc3-1, gmlERV-Fc-#1 (data not shown), gmlERV-Fc-#2, and oafERV-Fc1. Transfected cells were challenged with MLV particles pseudotyped with no Env, VSVG or babERV-Fc2-cl-MLVct (Fig. 3.14). When either babERV-Fc2 or babERV-Fc2-cl-MLVct was expressed in target cells, infection by MLV-babERV-Fc-2-cl-MLVct was significantly reduced. However, this was not observed with any of the other Envs (Fig.3.14). This indicates that babERV-Fc2, which is not cleaved by furin, is still able to interact with the putative receptor and block infection. Importantly, this is the protein encoded by the native locus in the baboon genome, suggesting that it could function as a restriction factor in baboon cells. The other ERV-Fc Envs failure to block infection could indicate either different receptor use, or that they are non-functional in some way.

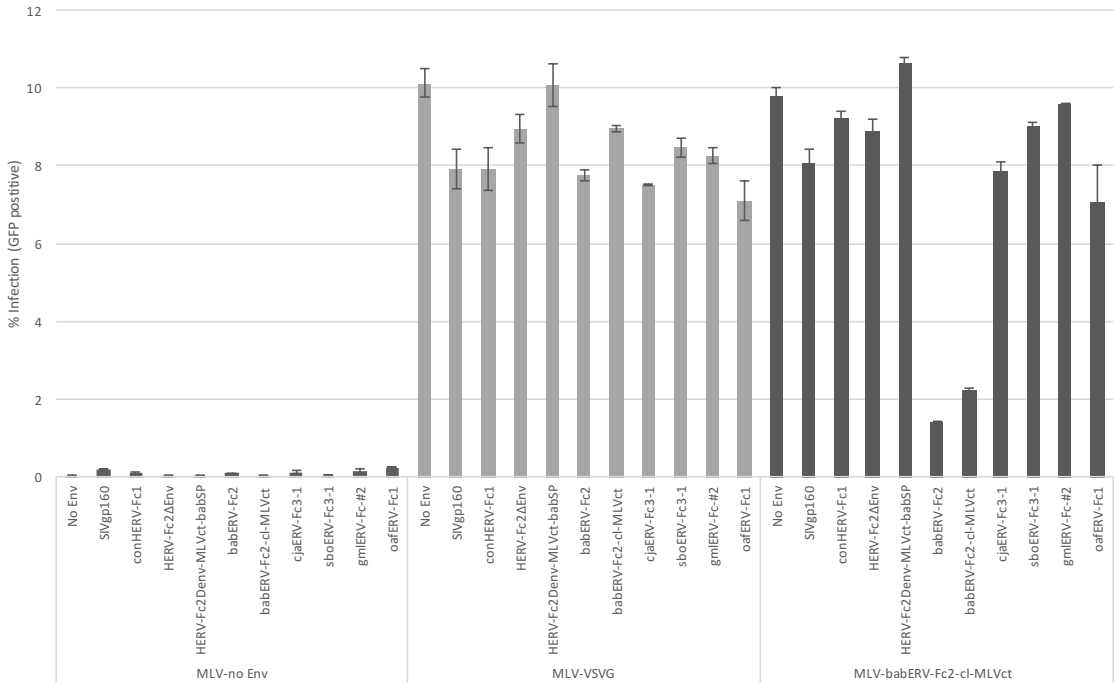


Figure 3.14: babERV-Fc2 Env blocks infection of MLV pseudotyped with babERV-Fc2-cl-MLVct. Infectivity of MLV particles pseudotyped with babERV-Fc2-cl-MLVct, VSVG or No Env when 293T/17 cells are expressing ERV-Fc Envs or control Env SIVgp160, N=2. The depicted experiment is a representative of 3-6 independent experiments.

The dog and panda ERV-Fc Envs are part of the RDR supergroup

The *envs* in this study were classical ERV-Fc-*envs* except for cfERV-Fc1(a) and ameERV-Fc1, which appear to have originated from a recombination event in the SU, resulting in an Env that branches with ERV-W in phylogenetic trees (Fig. 3.15A). Sequence analyses of cfERV-Fc1(a) and ameERV-Fc1 revealed the presence of the conserved ASCT2 interaction motif SDGGG2XD2XR and other conserved residues upstream of the motif (Fig. 3.15B). This suggests that cfERV-Fc1(a) and ameERV-Fc1 may (or did) use ASCT2 as a receptor, as seen with other RDR-supergroup Envs that have the motif [140]–[143]. When they are used as challenge Envs in a superinfection assay the cfERV-Fc1(a) and ameERV-Fc1 Envs are unable to block infection by MLV pseudotyped with babERV-Fc2-cl-MLVct (Fig. 3.16). This indicated that they may use a different receptor than babERV-Fc2, supporting cfERV-Fc1(a) and ameERV-Fc1s potential use of ASCT2. To test this possibility, we asked whether expression of cfERV-Fc1(a) and ameERV-Fc1 could block infection by MPMV, which is part of the RDR-supergroup and uses ASCT2 as a receptor. However, the results were inconclusive because the positive control (expression of the MPMV Env) did not work in the superinfection interference assay (Fig. 3.16). This may be due to transient transfections not being sufficient to block all the receptor present or other unknown factors. Infection data with cfERV-Fc1(a)-MLVct was also inconclusive, leading to little or no infection in 293T/17 cells (Fig. 3.13). This may be due to several factors, such as the Env no longer being functional, or cfERV-Fc1(a)'s inability to use human ASCT2 as a receptor.

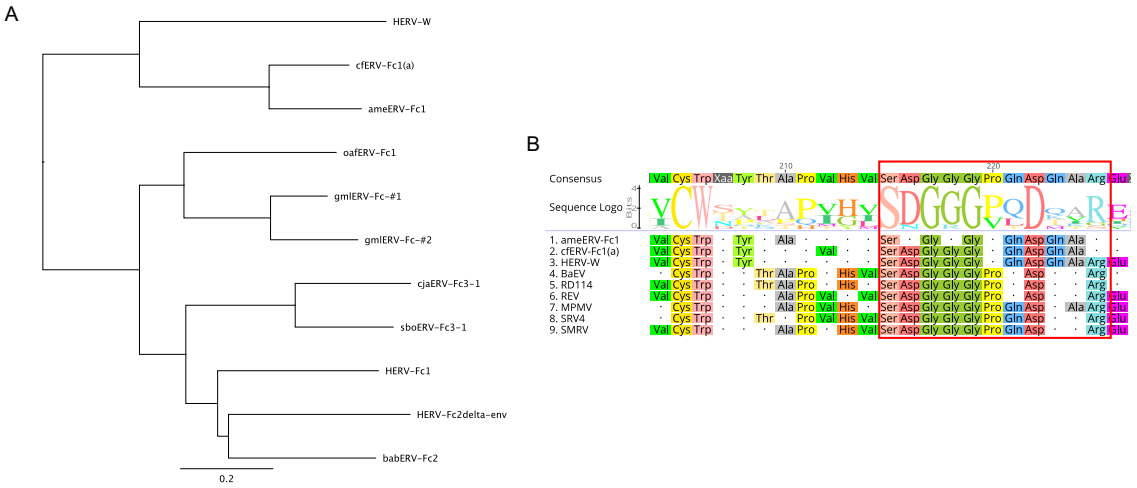


Figure 3.15: ameERV-Fc1 and cfERV-Fc1(a) SU domains have the conserved ASCT2 binding motif. (A) Neighbor-joining tree of full length ERV-Fc Envs and HERV-W. (B) Alignment of ameERV-Fc1 and cfERV-Fc1(a) with RDR supergroup Envs, red box highlights the conserved ASCT2 receptor binding motif (SDGGG2XD2XR).

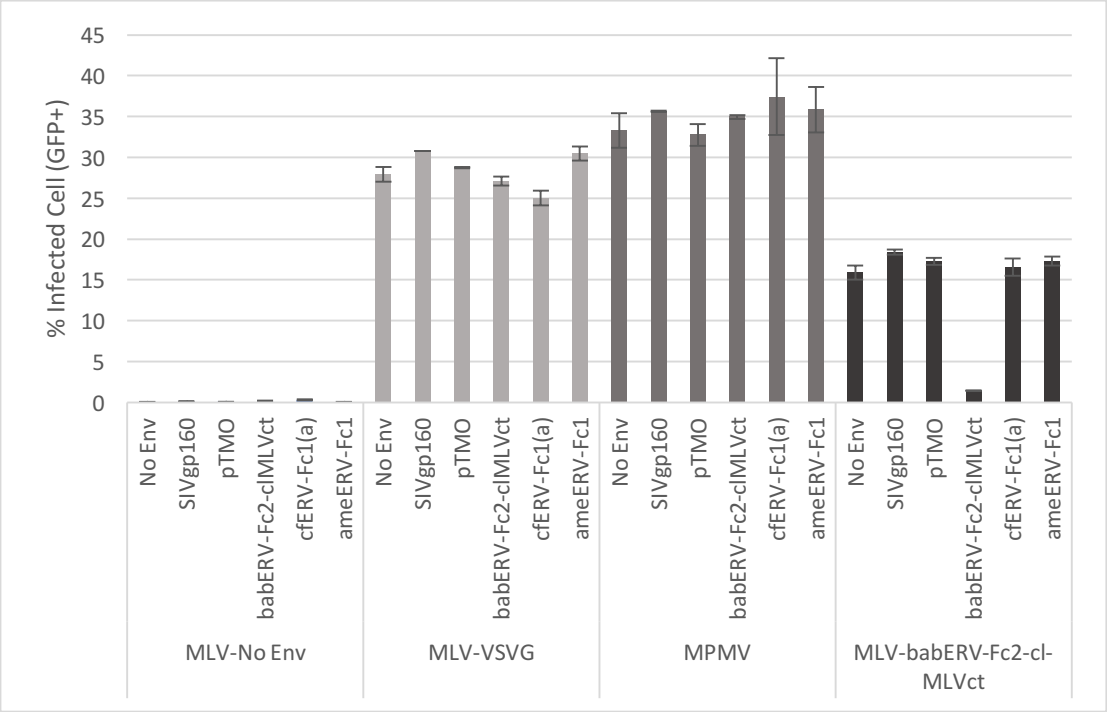


Figure 3.16: cfERV-Fc1(a) and ameERV-Fc1 do not block infection by virus pseudotyped with babERV-Fc2-cl-MLVct. Infectivity of MLV particles pseudotyped with babERV-Fc2-cl-MLVct, VSVG or No Env and MPMV when 293T/17 cells are expressing cfERV-Fc1(a) and ameERV-Fc1. Controls SIVgp160, pTMO, babERV-Fc2-cl-MLVct and No Env. N=2

Chapter 4: Discussion

Here, I have described the discovery and initial characterization of ten ERV-Fc *env* ORFs, which are found as part of partial or full proviruses in which the rest of the viral genes have acquired inactivating mutations (Fig. 3.1). The preservation of intact ORFs such as these is suggestive of exaptation for cellular/host function.

An alignment of the ERV-Fc Envs shows they have low identity and similarity in SU, but higher similarity and identity in TM (Fig. 3.3A and B). Codon optimized constructs of the *envs* were transfected into 293T/17 cells and protein expression was observed for all of them, though some expressed at higher levels than others (Fig. 3.4 A and B). However, most of the Env precursors were not processed properly between the SU and TM, even when there was an intact furin cleavage site. Additionally, HERV-Fc2∆env was not glycosylated (Fig. 3.4 A and C).

Several attempts were made to reconstitute functional versions of the human and baboon Envs. For example, adding the signal peptides from babERV-Fc2-env or conERV-Fc1-env to HERV-Fc2∆env restored glycosylation; however, repairing furin cleavage site mutants in HERV-Fc2∆env and conERV-Fc1 did not restore processing (Fig. 3.5, 3.6 and 3.7).

In the case of babERV-Fc2 Env, restoring a canonical furin cleavage site did restore SU-TM processing, but the Env was still not infectious in pseudotyped MLV assays (Fig. 3.8 and 3.10). This led to the hypothesis that ERV-Fc Envs may contain an R-peptide, similar to exogenous retroviruses like MLV and MPMV [18], [22], [79]. This was shown using truncation mutants and a chimeric Env

with the MLVct. Specifically, removing 22 amino acids from the babERV-Fc Env cytoplasmic tail led to the formation of syncytia, and infection was observed with this mutant and the chimeric Env (Fig. 3.9 and 3.10).

A screen of multiple cell lines indicates babERV-Fc2 utilizes a receptor present on human cell lines (Fig. 3.11 and 3.12). The identity of the receptor remains unknown. I made several limited attempts to clone the receptor by a gain-of-function screen of a 293T cell cDNA library using a transducible puromycin-resistance reporter (see Appendix xxx for a summary of the screen). Although a number of puro resistant colonies were obtained in a pilot screen, further testing revealed they were false positives likely due to a small background infection rate in CRFK cells. Given the small-scale nature of the screen, further efforts are warranted.

A superinfection interference assay revealed that the uncleaved version of babERV-Fc2 Env was able to block infection by virus pseudotyped with babERV-Fc2-cl-MLVct (Fig. 3.14). This confirms that the native babERV-Fc Env encoded in the baboon genome could function as an antiviral gene. A similar result was recently reported for an intact HERV-T Env ORF found in the human genome [134].

Finally, the dog and panda ERV-Fc Envs originated from a recombination event and resemble HERV-W (Fig. 3.15A). In the SU domains of these Envs, the proposed ASCT2 receptor interaction motif SDGGGX2DX2R is present, suggesting they are part of the RDR supergroup of Envs (3.15B). I was unable to generate infectious pseudotypes with either Env, so it was not possible to

confirm that cfERV-Fc1(a) and ameERV-Fc1 use ASCT2 as a receptor directly. In the future, assays based on direct binding, such as Co-IP, may be necessary.

The data reported suggest that the ERV-Fc-*env* ORFs may have been conserved by selection for a host function. While each of the *env* ORFs are intact, in each case the rest of the ERV-Fc provirus is mutated and the other ORFs are disrupted, suggesting there was selective pressure to maintain the *env* ORFs. All the codon optimized constructs were expressed in transfected cells; however, there were multiple defects affecting cleavage at the furin cleavage site between the SU and TM, indicating that cleavage and fusion were not maintained by selection. This is supported by the presence of a possible R-peptide in babERV-Fc2, which can be assumed to be present on all the ERV-Fc Envs due to sequence similarity at the site. Interestingly, the possible R-peptide in the ERV-Fcs does not appear to be cleaved by MLV protease. It would be interesting to ask whether there is incompatibility between gammaretroviral Env R-peptides, since protease is highly conserved in retroviruses. The known MLV R-peptide cleavage site (L-V) is not present in the ERV-Fc Env cytoplasmic tail. This raises several possibilities, first the babERV-Fc2-env lost the R-peptide cleavage site used by MLV and several other Envs [82]. However, the cytoplasmic tail of the ERV-Fcs around the predicted area of the R-peptide cleavage site is highly conserved and it is unlikely that all the ERV-Fc Envs lost a potential cleavage site in the same manner. As a result, we can speculate that the MLV protease does not recognize the cleavage site in the ERV-Fc protease. The incompatibility may be due to the age of ERV-Fc (~33 to 15 million years old), during that time there

may have been slight changes in the viral protease. Alternately, ERV-Fc Envs may be uniquely adapted to their own protease. The MPMV R-peptide cleavage site is between a Y and H, a different site than MLV R-peptide cleavage site, which may suggest that the MPMV R-peptide cleavage site is specific to its own protease [81]. Thus, a similar adaptation may have occurred in ERV-Fc. Some ERV Envs are fusogenic, without requiring cleavage of an R-peptide. An example being HERV-W/*syncytin-1,* which is fusogenic without cleaving off an R-peptide, and which does not have the MLV or MPMV R-peptide cleavage site in its cytoplasmic tail [90], [144], [145]. However, ERV-Fc does not seem to have this ability, as evidenced by babERV-Fc-cl's inability to mediate fusion. Currently it is unclear why the MLV protease cannot cleave off the ERV-Fc R-peptide; however, because of this incompatibility it suggests that ERV-Fc may have been fusion incompetent when it first integrated in the baboon genome.

This fusion deficiency suggests a reason for maintaining the ERV-Fc *env* ORFs. Cell-cell fusion is unable to occur without a truncated cytoplasmic tail and may be damaging to the host, and an ERV Env that can cause this reaction will likely be selected against (unless its function involves fusion, as with the *syncytins*). The fusion deficiency in several of the ERV-Fc Envs was compounded by multiple other mutations that interfered with the furin cleavage site, deactivating cleavage and further preventing fusion. The trend of ERV-Env's being defective, especially fusion defective, has been observed among other ERV Envs that have been implicated in receptor interference antiviral activity. For example, Refrex-1 in cats is truncated before the MSD, but still able

to bind to a receptor and block it from use by an exogenous virus when it is secreted from the cell [114]. Additionally, when the ancestral Refrex-1 was reconstructed there were mutations upstream of the furin cleavage site that blocked processing between the SU and TM [131]. This raises the possibility that it was also unable to induce cell-cell fusion before it was truncated. We observed the same pattern of furin cleavage inactivation in several of the ERV-Fc Envs, where three of the Envs (conERV-Fc1, cjaERV-Fc3 and gmlERV-Fc#1) have canonical furin cleavage sites; however, they are not properly processed, suggesting other mutations distal to the furin target site prevent cleavage. These mutations likely block by changing the protein structure. Similar to Refrex-1, the mostly open HERV-T Env has mutations that deactivate the furin cleavage site, both in and outside the site [134]. Fv4 in mice has a defective fusion peptide and cannot mediate viral entry into a cell [126]. All these Envs are fusion defective as their furin cleavage sites are inactivated, their fusion peptide is defective, or in the case of ERV-Fc, they have an R-peptide that has to be cleaved off. I propose that inactivation of the fusogenic ability of an ERV Env by some or all of these mechanisms may be a hallmark of exaptation, particularly for use in an antiviral function involving receptor interference.

There are multiple examples of ERV Envs being co-opted for an antiviral function; which may include the ERV-Fc Envs. This is supported by the ability of the native babERV-Fc2 Env (without the reconstructed furin cleavage site and with the R-peptide) to block infection by pseudotyped virus carrying the babERV-Fc2-ENV-cl-MLVct. These results indicate that cleavage at both places is not

necessary for receptor binding and is sufficient to block infection. This observation was also seen with HERV-T, HERV-Ts potentially antiviral Env is not cleaved, but it is still able to block infection from an exogenous virus [134]. The other ERV-Fc Envs were not able to block infection due to several factors: they may not be functional, a human receptor is not used, or they may utilize a different receptor than babERV-Fc2. All have very low identity and similarity, especially in the SU, which is responsible for receptor recognition. This suggests that the ERV-Fc Envs may use several different receptors. We know that ERV-Fc uses at least two different receptors, one that babERV-Fc2 recognizes and another for cfERV-Fc1(a) and ameERV-Fc1. Their Envs are the result of a recombination event and have similarities to Envs from the RDR supergroup, including a known receptor interaction motif SDGGGX2DX2R which is not present in the other ERV-Fc Envs. We speculate that endogenization may even drive the use of different receptors. Ecotropic, xenotropic, polytrophic and amphotropic strains of MLV use of distinct receptors, CAT-1, Rmc1 and Ram-1 respectively, even though they are closely related, with the change in receptor being caused by only a few mutations [1], [17]. Fv4 is only capable of blocking infection by ecotropic MLV [115], [146]. Rmfc1 and Rmfc2 are only able to block infection by polytrophic MLV [116], [117]. KoRV-A, is currently in the processes of endogenizing in koalas, and uses PiT1 as a receptor [147]. A recent second form KoRV-B is not endogenous and uses THTR1 as a receptor even though it is closely related to KoRV-A [148]–[150]. The ability of these ERV Envs to block only one form of the virus and the fact that they use different receptors from

closely related viruses suggests there may have been selection to change receptor usage. The sole purpose of a virus is to infect a host replicate and continue to spread. Therefore, if a receptor is blocked from use, selective pressure may drive the use of a new receptor so that the virus continues to replicate. This may be what happened with MLV and KoRV. It could also explain why the ERV-Fc Env SUs are divergent from each other, perhaps as ERV-Fc started to endogenize the exogenous form of ERV-Fc adapted to move around the receptor block adapting to use multiple different receptors. This suggests there may have previously been more ERV-Fc *env* ORFs.

Our results suggest the following scenario for the endogenization and exaptation of antiviral ERV Envs. As an exogenous retrovirus spreads within a species, the acquisition of germline insertions results in ERVs that express Envs (Fig. 4.1). Some of these ERV Envs are able to block the receptor used by the exogenous virus protecting the host. Individuals carrying these ERV-Fc insertions may have gained a fitness advantage in the form of viral resistance, leading to fixation of the protective alleles (Fig. 4.1). This selection for the fixation of the ERV is driven by the still spreading exogenous virus. During this time, additional modifications to the ERV Env in order to preserve receptor binding function and/or to eliminate fusogenicity may have also been selected for, as I discovered in the ERV-Fc Envs, and which have been reported for Fv-4, Refrex-1 and HERV-T. In contrast to antiviral ERV Envs, the *syncytins* retain fusion ability [87]. They play an important role in pregnancy; however, having fusogenic protein constitutively expressed could be harmful. *Syncytins* must be

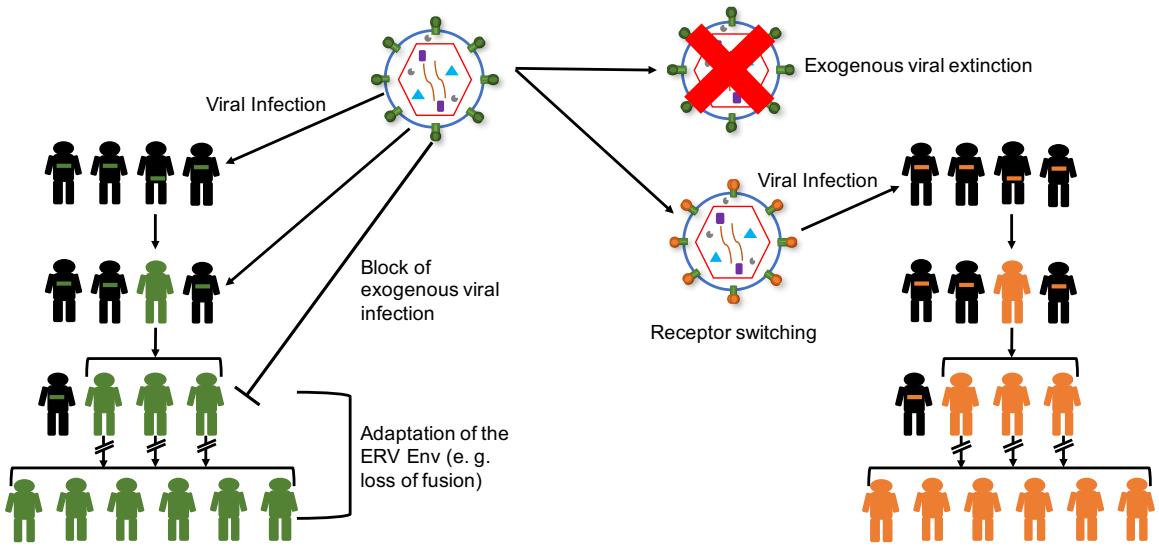


Figure 4.1: Endogenization and exaptation of antiviral ERV Envs. On the left, spread of an exogenous retrovirus results in an infected germline cell and an endogenization event (left green). As the exogenous virus continues to spread it selects for individuals with the ERV who are protected from the virus. This leads to fixation of the ERV within a population. Pressure from the receptor being blocked results in exogenous virus extinction (top right) or a receptor switch allowing the exogenous virus to continue spreading in a population (bottom right) and may result in a new endogenization event (right orange).

heavily regulated and expressed only during certain developmental stages and in specific tissues [87]. The antiviral ERVs may still be regulated and only expressed in certain tissue types but they may be less harmful to the host if they are expressed at the wrong time or place due to many of them being fusion deficient.

The fixation of an antiviral ERV may eventually cause the extinction of the exogenous virus from the population (Fig. 4.1). Alternatively, fixation of the protective ERV Env in the host population could drive the exogenous virus to adapt by switching receptors. The receptor switch can be achieved either through substitutions in the SU, or by recombination and replacement of the receptor binding domains with the receptor binding domains of retroviruses that use a different receptor, thereby allowing the exogenous virus to continue spreading within a population (Fig. 4.1). The change in the Env leading to a receptor switch may also result in infection in a new species. If another germline infection occurs leading to an ERV this model of the establishment of an antiviral ERV Env may repeat itself on evolutionary timescales (Fig. 4.1). The ten loci ERV-Fc Envs from eight vertebrate lineages described is this work may represent similar but independent occurrences of this endogenization and exaptation scenario. Taken together with examples from mice (*Fv-4, rmfc-1* and *2*), chickens (*ev3, 6* and *9*) and cats (Refrex-1), my results suggest that long-term propagation of gammaretroviruses in a population selects for the appearance and exaptation of protective ERV loci.

I utilized multiple Envs from different species to form a more complete picture of what the endogenization process of a family of ERVs may look like, and by taking into account our results and previous work done with ERV Envs, the above scenario was established. ERV-Fc is widespread throughout mammals, spreading for many millions of years and the model we propose may help explain why that is [77]. As discussed above, ERV-Fc Envs are non-fusogenic from the moment they were endogenized, automatically making them good candidates for an antiviral function. An envelope capable of fusion could be harmful for the host and may not be maintained. A non-fusogenic Env that can protect a host from infection may be maintained, and additional mutations that further suppress fusion could subsequently be selected (as seen in the ERV-Fc Envs and other antiviral ERV Envs). The second factor that may have contributed to the widespread occurrence of ERV-Fc may be due to receptor switching, in turn the result of pressure from receptor blocking by an endogenous Env, which have broadened the tropism of the ERV (Fig. 4.1). Recombination, as with the ERV-Fc Env in the dog and panda genomes, changed ERV-Fcs receptor usage. Low identity in the SU also suggests the use of multiple receptors. Both of these could lead to the exogenous ERV-Fc crossing over to infect multiple different species resulting in the wide spread ERV-Fc fossils we now see.

While there are currently only examples of ERV-Fc-*env* ORFs in 10 species, ERV-Fc is so widespread in mammals we hypothesize that historically there were more *env* ORFs that were only transiently open to perform a function

when needed, for example, to block a viral receptor. However, if the exogenous form of ERV-Fc adapted to use a new receptor, there would no longer be pressure to maintain the existing *env* ORF. We propose the ERV-Fc *env* ORFs were maintained to act as antiviral factors through receptor interference. Although there are documented examples of ERV Envs acting as antiviral entry inhibitors, this is the first study to couple repeated exaptation of *env* genes to the long-term evolution of a specific viral lineage (ERV-Fc). Future work focusing on identifying the receptor(s) for these Envs could shed light on the exaptive process and its correlation with viral extinction and the evolution of receptor-usage/receptor-switching.

Appendix 1: cDNA Receptor Screen

Materials and Methods

cDNA Library

A cDNA library was created from 293T cells using the SMART cDNA Library construction kit by Ismael Fofana (Clontech). cDNA fragments were cloned into the pRETRO-Lib vector between the Sfi IA and Sfi IB sites.

Cell Lines and Media

Cell lines (HEK293T/17, CRFK) were obtained from the American Type Culture Collection (ATCC). Cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM). Media was supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine (200mM), 1% Penicillin-Streptomycin (10,000 IU-10,000 µg/ml) and 2.5% HEPES (1M). Cells were incubated at 37°C.

Transduction of cDNA Library into CRFK Cells

In order to pseudotyped MLV particles caring the 293T cDNA library, 293T/17 cells seeded in a 6 well plate for a confluence of ~70%, were transfected with pRETRO-Lib, pCIGB and a plasmid containing VSV-G. The plasmids were transfected at a ratio of 3:2:1 respectively using GenJet (SignaGen® Laboratories), resulting in a total DNA concentration of 2µg/well.

48 hours after transfection, supernatant was harvested from the 293T/17 cells centrifuged at 5000rpm for 5 minutes. CRFK cells seeded in a 12 well plate at ~50% confluency were transduced with 250µl of supernatant containing cDNA viral particles.

cDNA Library Screen

Two days after transduction, CRFK cells were screened with MLV particles pseudotyped with babERV-Fc2-cl-MLVct and carrying the *puro* resistance gene (pBABE-puro Addgene). 48 hours after the cDNA containing CRFK cells were challenged, they were put under puromycin selection (10μg/μl). Cells were kept under selection until puro resistant colonies appeared. Colonies were placed in new plates. Two subsequent assays were done with puro resistant colonies. First, they were challenged with MLV particles pseudotyped with babERV-Fc2-cl-MLVct and containing a GFP reporter gene. GFP positive cells were then counted using Flow. Second, the genomic DNA from puro resistant colonies was harvested using a genomic DNA harvesting kit (Promega). The cDNA fragments were then PCR amplified using primers specific to the flanking region (Clonetech). PCR fragments were then gel extracted and blunt end cloned into a TOPO vector (ThermoFisher Scientific). Clones were then sent for sequencing (Eton Biosciences).

Results

In order to determine the receptor used by babERV-Fc2 a cDNA screen was performed. A cDNA library was made using mRNA extracted from 293T cells, because the highest infection with babERV-Fc2 Env-pseudotyped virus was seen in those cells (Fig. 3.11). The cDNA library was then transduced into CRFK cells because babERV-Fc2 is unable to infect CRFK cells and there is low background (Fig. 3.11). CRFK cells transduced with the 293T cDNA library were challenged with MLV particles pseudotyped with babERV-Fc2-cl-MLVct carrying the *puro* resistant gene. Puromycin selection was then used to screen for CRFK cells that were then susceptible to infection by MLV particles pseudotyped with babERV-Fc2-cl-MLVct. Five separate cDNA screens were done. From those screens 25 resistant colonies were isolated and further tested (Fig. A1.1). Resistant clones were reinfected by MLV-babERV-Fc2-cl-MLVct, however there was not a significant improvement in susceptibility to infection when compared to CRFK cells without the 293T cDNA library (Fig. A1.1). This indicates that cDNAs present in these puro resistant CRFK colonies did not encode the receptor used by babERV-Fc2, and are most likely the result of background infection. This is supported by the sequencing results from the puro resistant colonies. While partial gene sequences were obtained, none were good receptor candidates (Fig.A1.1). Gamma Envs are known to used membrane proteins that have multiple membrane spanning domains and are often transporters [17]. Further larger scale screens will need to be done in order to determine the receptor used by babERV-Fc2.

A

Colony	Sequencing Results
2C1	Vector
2C3	histidine triad nucleotide binding protein 1
7C3	None
8B1	Actin regulator (ENAH)
9A2	nucleophosmin (nucleolar phosphoprotein B23, numatin), and FLJ95438, cell division cycle 40 homolog
9B3	Vector
9C4	cyclin B1 interacting protein 1, and Thioredoxin
10B2	ATP synthase, H+ trasporting , mitochondrial Fo complex, subunit F6
10B3	CGG triplet repeate binding protein 1
10B4	actin regulator (ENAH)
10C4	metabolism of cobalamin associated (MMACHC) trascript varient 1
11A4	Ribosomal protein S4
11B3	Ribosomal protein S4
12B3	Ribosomal protein S4
12B4	Ribosomal protein S4
13B3	None
14B1	None
15A1-1	None
15A1-2	None
15A2	None
19A3	None
22B1	None
22B2-1	None
22B2-2	None
27A2	None

B

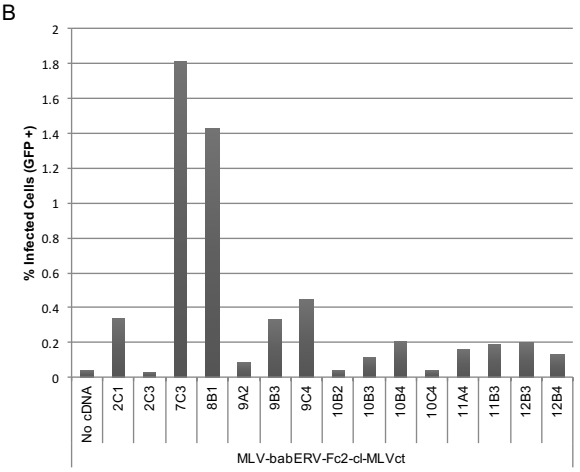


Figure A1.1: cDNA library screen results. (A) A list of all puromycin resistant colonies, named for the plate#:row:well# they were found in, listed also is the sequencing results obtained from each colony. (B) Reinfection of puromycin resistant colonies with MLV-babERV-Fc2-cl-MLVct.

References

[1] B. N. Fields, D. M. Knipe, and P. M. Howley, *Fields virology*. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2007.

[2] J. Coffin, S. Hughes, and H. Varmus, *Retroviruses*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Pres, 1997.

[3] B. Bowerman, P. O. Brown, J. M. Bishop, and H. E. Varmus, "A nucleoprotein complex mediates the integration of retroviral DNA.," *Genes Dev.*, vol. 3, no. 4, pp. 469–478, Apr. 1989.

[4] H. M. Temin and S. Mizutani, "RNA-dependent DNA polymerase in virions of Rous sarcoma virus. 1970.," *Biotechnology*, vol. 24, pp. 51–56, 1992.

[5] D. Baltimore, "RNA-dependent DNA polymerase in virions of RNA tumour viruses.," *Nature*, vol. 226, no. 5252, pp. 1209–1211, Jun. 1970.

[6] P. F. Lewis and M. Emerman, "Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus.," *J. Virol.*, vol. 68, no. 1, pp. 510–516, Jan. 1994.

[7] E. H. Humphries and H. M. Temin, "Cell cycle-dependent activation of rous sarcoma virus-infected stationary chicken cells: avian leukosis virus group-specific antigens and ribonucleic acid.," *J. Virol.*, vol. 10, no. 1, pp. 82–87, Jul. 1972.

[8] E. H. Humphries, C. Glover, R. A. Weiss, and J. R. Arrand, "Differences between the endogenous and exogenous DNA sequences of Rous-associated virus-O.," *Cell*, vol. 18, no. 3, pp. 803–815, Nov. 1979.

[9] D. G. Miller, M. A. Adam, and A. D. Miller, "Gene transfer by retrovirus vectors occurs only in cells that are actively replicating at the time of infection.," *Mol. Cell. Biol.*, vol. 10, no. 8, pp. 4239–4242, Aug. 1990.

[10] A. T. Haase, "The slow infection caused by visna virus.," *Curr. Top. Microbiol. Immunol.*, vol. 72, pp. 101–156, 1975.

[11] M. S. Meltzer, D. R. Skillman, P. J. Gomatos, D. C. Kalter, and H. E. Gendelman, "Role of mononuclear phagocytes in the pathogenesis of human immunodeficiency virus infection.," *Annu. Rev. Immunol.*, vol. 8, pp. 169–194, 1990.

[12] A. M. Fadly, "Avian retroviruses.," *Vet. Clin. North Am. Food Anim. Pract.*, vol. 13, no. 1, pp. 71–85, Mar. 1997.

[13] J. E. Henzy and W. E. Johnson, "Pushing the endogenous envelope," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 368, no. 1626, pp. 20120506–20120506, 2013.

[14] S. R. Ross, "Mouse mammary tumor virus molecular biology and oncogenesis.," *Viruses*, vol. 2, no. 9, pp. 2000–2012, Sep. 2010.

[15] C. Leroux, N. Girard, V. Cottin, T. Greenland, J.-F. Mornex, and F. Archer, "Jaagsiekte Sheep Retrovirus (JSRV): from virus to lung cancer in sheep.," *Vet. Res.*, vol. 38, no. 2, pp. 211–228, 2007.

[16] M. Brack, "[Simian AIDS virus: review].," *Zentralbl. Veterinarmed. B*, vol. 36, no. 10, pp. 721–745, Dec. 1989.

[17] J. Overbaugh, A. D. Miller, and M. V Eiden, "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.*, vol. 65, no. 3, p. 371–89–table of contents, Sep. 2001.

[18] W. L. Karshin, L. J. Arcement, R. B. Naso, and R. B. Arlinghaus, "Common Precursor for Rauscher Leukemia Virus gp69 / 71 , p15 (E), and p12 (E)," *J. Virol.*, vol. 23, no. 3, pp. 787–798, 1977.

[19] J. G. Sutcliffe, T. M. Shinnick, N. Green, F. T. Liu, H. L. Niman, and R. A. Lerner, "Chemical synthesis of a polypeptide predicted from nucleotide sequence allows detection of a new retroviral gene product.," *Nature*, vol. 287, no. 5785, pp. 801–805, Oct. 1980.

[20] N. Green, T. M. Shinnick, O. Wittet, A. Ponticellit, J. G. Sutcliffe, and R. A. Lerner, "Sequence-specific antibodies show that maturation of Moloney leukemia virus envelope polyprotein involves removal of a COOH-terminal peptide," *Proc. Natl. Acad. Sci. USA*, vol. 78, no. 10, pp. 6023–6027, 1981.

[21] L. E. Henderson, R. Sowder, T. D. Copeland, G. Smythers, and S. Oroszlan, "Quantitative Separation of Murine Leukemia Virus Proteins by Reversed-Phase High-Pressure Liquid Chromatography Reveals Newly Described gag and env Cleavage Products," *J. Virol.*, vol. 52, no. 2, pp. 492–500, 1984.

[22] A. Rein, J. Mirro, J. G. Haynes, S. M. Ernst, and K. Nagashima, "Function of the Cytoplasmic Domain of a Retroviral Transmembrane Protein : pl5E-p2E Cleavage Activates the Membrane Fusion Capability of the Murine Leukemia Virus Env Protein," *J. Virol.*, vol. 68, no. 3, pp. 1773–1781, 1994.

[23] J. A. Ragheb and W. F. Anderson, "pH-Independent Murine Leukemia Virus Ecotropic Envelope- Mediated Cell Fusion : Implications for the Role of the R Peptide and pl2E TM in Viral Entry," *J. Virol.*, vol. 68, no. 5, pp. 3220–3231, 1994.

[24] A. Rein, "Murine leukemia viruses: objects and organisms.," *Adv. Virol.*, vol. 2011, p. 403419, 2011.

[25] B. J. Willett and M. J. Hosie, "Feline leukaemia virus: half a century since its discovery.," *Vet. J.*, vol. 195, no. 1, pp. 16–23, Jan. 2013.

[26] J. Denner and P. R. Young, "Koala retroviruses: characterization and impact on the life of koalas.," *Retrovirology*, vol. 10, p. 108, Oct. 2013.

[27] U. N. Mui, C. T. Haley, and S. K. Tyring, "Viral Oncology: Molecular Biology and Pathogenesis.," *J. Clin. Med.*, vol. 6, no. 12, Nov. 2017.

[28] I. Schwartz and D. Levy, "Pathobiology of bovine leukemia virus.," *Vet. Res.*, vol. 25, no. 6, pp. 521–536, 1994.

[29] D. L. Holzschu, D. Martineau, S. K. Fodor, V. M. Vogt, P. R. Bowser, and J. W. Casey, "Nucleotide sequence and protein analysis of a complex piscine retrovirus, walleye dermal sarcoma virus.," *J. Virol.*, vol. 69, no. 9, pp. 5320–5331, Sep. 1995.

[30] Z. Zhang, D. Du Tremblay, B. F. Lang, and D. Martineau, "Phylogenetic and epidemiologic analysis of the walleye dermal sarcoma virus.," *Virology*, vol. 225, no. 2, pp. 406–412, Nov. 1996.

[31] L. A. LaPierre, D. L. Holzschu, G. A. Wooster, P. R. Bowser, and J. W. Casey, "Two closely related but distinct retroviruses are associated with

walleye discrete epidermal hyperplasia.," *J. Virol.*, vol. 72, no. 4, pp. 3484–3490, Apr. 1998.
[32] R. F. Cook, C. Leroux, and C. J. Issel, "Equine infectious anemia and equine infectious anemia virus in 2013: a review.," *Vet. Microbiol.*, vol. 167, no. 1–2, pp. 181–204, Nov. 2013.
[33] M. L. Linial, "Foamy viruses are unconventional retroviruses.," *J. Virol.*, vol. 73, no. 3, pp. 1747–1755, Mar. 1999.
[34] J. P. Stoye, "Studies of endogenous retroviruses reveal a continuing evolutionary saga," *Nat. Rev. Microbiol.*, vol. 10, p. 395, May 2012.
[35] F. Benachenhou *et al.*, "Evolutionary conservation of orthoretroviral long terminal repeats (LTRs) and ab initio detection of single LTRs in genomic data.," *PLoS One*, vol. 4, no. 4, p. e5179, 2009.
[36] N. G. Copeland, K. W. Hutchison, and N. A. Jenkins, "Excision of the DBA ecotropic provirus in dilute coat-color revertants of mice occurs by homologous recombination involving the viral LTRs.," *Cell*, vol. 33, no. 2, pp. 379–387, Jun. 1983.
[37] W. E. Johnson and J. M. Coffin, "Constructing primate phylogenies from ancient retrovirus sequences.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 18, pp. 10254–10260, Aug. 1999.
[38] A. Shih, E. E. Coutavas, and M. G. Rush, "Evolutionary implications of primate endogenous retroviruses.," *Virology*, vol. 182, no. 2, pp. 495–502, Jun. 1991.
[39] A. W. Dangel, B. J. Baker, A. R. Mendoza, and C. Y. Yu, "Complement component C4 gene intron 9 as a phylogenetic marker for primates: long terminal repeats of the endogenous retrovirus ERV-K(C4) are a molecular clock of evolution.," *Immunogenetics*, vol. 42, no. 1, pp. 41–52, 1995.
[40] D. L. Mager and J. D. Freeman, "HERV-H endogenous retroviruses: presence in the New World branch but amplification in the Old World primate lineage.," *Virology*, vol. 213, no. 2, pp. 395–404, Nov. 1995.
[41] P. SanMiguel, B. S. Gaut, A. Tikhonov, Y. Nakajima, and J. L. Bennetzen, "The paleontology of intergene retrotransposons of maize.," *Nat. Genet.*, vol. 20, no. 1, pp. 43–45, Sep. 1998.
[42] J. D. Boeke and J. P. Stoye, "Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements," J. M. Coffin, S. H. Hughes, and H. E. Varmus, Eds. Cold Spring Harbor (NY), 1997.
[43] I. A. Maksakova, D. L. Mager, and D. Reiss, "Keeping active endogenous retroviral-like elements in check: the epigenetic perspective.," *Cell. Mol. Life Sci.*, vol. 65, no. 21, pp. 3329–3347, Nov. 2008.
[44] D. C. Leung and M. C. Lorincz, "Silencing of endogenous retroviruses: when and why do histone marks predominate?," *Trends Biochem. Sci.*, vol. 37, no. 4, pp. 127–133, Apr. 2012.
[45] H. M. Rowe and D. Trono, "Dynamic control of endogenous retroviruses during development.," *Virology*, vol. 411, no. 2, pp. 273–287, Mar. 2011.
[46] R. A. Weiss, "The discovery of endogenous retroviruses," *Retrovirology*, vol. 3, no. 1, p. 67, 2006.
[47] R. M. Dougherty and H. S. Di Stefano, "Lack of relationship between

infection with avian leukosis virus and the presence of COFAL antigen in chick embryos.," *Virology*, vol. 29, no. 4, pp. 586–595, Aug. 1966.
[48] R. M. Dougherty, H. S. Di Stefano, and F. K. Roth, "Virus particles and viral antigens in chicken tissues free of infectious avian leukosis virus.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 58, no. 3, pp. 808–817, Sep. 1967.
[49] L. N. Payne and R. C. Chubb, "Studies on the nature and genetic control of an antigen in normal chick embryos which reacts in the COFAL test.," *J. Gen. Virol.*, vol. 3, no. 3, pp. 379–391, Dec. 1968.
[50] R. Weiss, "Spontaneous virus production from 'non-virus producing' Rous sarcoma cells.," *Virology*, vol. 32, no. 4, pp. 719–723, Aug. 1967.
[51] P. K. Vogt, "A virus released by 'nonproducing' Rous sarcoma cells.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 58, no. 3, pp. 801–808, Sep. 1967.
[52] R. Weiss, "The Host Range of BRYAN Strain Rous Sarcoma Virus Synthesized in the Absence of Helper Virus," *J. Gen. Virol.*, vol. 5, pp. 511–528, Dec. 1969.
[53] R. A. Weiss, "Interference and neutralization studies with Bryan strain Rous sarcoma virus synthesized in the absence of helper virus," *Journal of general virology*. 1969.
[54] R. A. Weiss and L. N. Payne, "The heritable nature of the factor in chicken cells which acts as a helper virus for Rous sarcoma virus.," *Virology*, vol. 45, no. 2, pp. 508–515, Aug. 1971.
[55] R. A. Weiss, R. R. Friis, E. Katz, and P. K. Vogt, "Induction of avian tumor viruses in normal cells by physical and chemical carcinogens.," *Virology*, vol. 46, no. 3, pp. 920–938, Dec. 1971.
[56] S. M. Astrin, H. L. Robinson, L. B. Crittenden, E. G. Buss, J. Wyban, and W. S. Hayward, "Ten genetic loci in the chicken that contain structural genes for endogenous avian leukosis viruses.," *Cold Spring Harb. Symp. Quant. Biol.*, vol. 44 Pt 2, pp. 1105–1109, 1980.
[57] P. N. Rosenthal, H. L. Robinson, W. S. Robinson, T. Hanafusa, and H. Hanafusa, "DNA in uninfected and virus-infected cells complementary to avian tumor virus RNA.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 68, no. 10, pp. 2336–2340, Oct. 1971.
[58] H. E. Varmus, R. A. Weiss, R. R. Friis, W. Levinson, and J. M. Bishop, "Detection of avian tumor virus-specific nucleotide sequences in avian cell DNAs (reassociation kinetics-RNA tumor viruses-gas antigen-Rous sarcoma virus, chick cells).," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 69, no. 1, pp. 20–24, Jan. 1972.
[59] M. A. Baluda, "Widespread presence, in chickens, of DNA complementary to the RNA genome of avian leukosis viruses.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 69, no. 3, pp. 576–580, Mar. 1972.
[60] W. P. Rowe and T. Pincus, "Quantitative studies of naturally occurring murine leukemia virus infection of AKR mice.," *J. Exp. Med.*, vol. 135, no. 2, pp. 429–436, Feb. 1972.
[61] J. P. Stoye, "Fv1, the mouse retrovirus resistance gene.," *Rev. Sci. Tech.*, vol. 17, no. 1, pp. 269–277, Apr. 1998.
[62] D. R. Lowy, W. P. Rowe, N. Teich, and J. W. Hartley, "Murine leukemia

virus: high-frequency activation in vitro by 5-iododeoxyuridine and 5-bromodeoxyuridine.," *Science*, vol. 174, no. 4005, pp. 155–156, Oct. 1971.
[63] S. A. Aaronson, G. J. Todaro, and E. M. Scolnick, "Induction of murine C-type viruses from clonal lines of virus-free BALB-3T3 cells.," *Science*, vol. 174, no. 4005, pp. 157–159, Oct. 1971.
[64] S. A. Aaronson, J. W. Hartley, and G. J. Todaro, "Mouse leukemia virus: 'spontaneous' release by mouse embryo cells after long-term in vitro cultivation.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 64, no. 1, pp. 87–94, Sep. 1969.
[65] M. A. Martin, T. Bryan, S. Rasheedt, and A. S. Khan, "Identification and cloning of endogenous retroviral , sequences present in human DNA Biochemistry :," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 78, no. 8, pp. 4892–4896, 1981.
[66] P. Medstrand and J. Blomberg, "Characterization of Novel Reverse Transcriptase Encoding Human Endogenous Retroviral Sequences Similar to Type A and Type B Retroviruses : Differential Transcription in Normal Human Tissues," *J. Virol.*, vol. 67, no. 11, pp. 6778–6787, 1993.
[67] E. S. Lander *et al.*, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.
[68] R. H. Mouse Genome Sequencing Consortium, Waterston *et al.*, "Initial sequencing and comparative analysis of the mouse genome.," *Nature*, vol. 420, no. 6915, pp. 520–562, Dec. 2002.
[69] International Chicken Genome Sequencing Consortium, "Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.," *Nature*, vol. 432, no. 7018, pp. 695–716, Dec. 2004.
[70] J. F. Hughes and J. M. Coffin, "Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: Implications for human and viral evolution," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 6, p. 1668 LP-1672, Feb. 2004.
[71] R. Belshaw, A. L. A. Dawson, J. Woolven-allen, J. Redding, A. Burt, and M. Tristem, "Genomewide Screening Reveals High Levels of Insertional Polymorphism in the Human Endogenous Retrovirus Family HERV-K (HML2): Implications for Present-Day Activity," *J. Virol.*, vol. 79, no. 19, pp. 12507–12514, 2005.
[72] J. H. Wildschutte, Z. H. Williams, M. Montesion, R. P. Subramanian, J. M. Kidd, and J. M. Coffin, "Discovery of unfixed endogenous retrovirus insertions in diverse human populations.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 16, pp. E2326-34, Apr. 2016.
[73] R. E. Tarlinton, J. Meers, and P. R. Young, "Retroviral invasion of the koala genome.," *Nature*, vol. 442, no. 7098, pp. 79–81, Jul. 2006.
[74] L. Bénit, A. Calteau, and T. Heidmann, "Characterization of the low-copy HERV-Fc family: evidence for recent integrations in primates of elements with coding envelope genes," *Virology*, vol. 312, no. 1, pp. 159–168, Jul. 2003.
[75] L. Bénit, P. Dessen, and T. Heidmann, "Identification, phylogeny, and

evolution of retroviral elements based on their envelope genes.," *J. Virol.*, vol. 75, no. 23, pp. 11709–11719, Dec. 2001.

[76] N. De Parseval, V. Lazar, J.-F. Casella, L. Benit, and T. Heidmann, "Survey of Human Genes of Retroviral Origin : Identification and Transcriptome of the Genes with Coding Capacity for Complete Envelope Proteins," *J. Virol.*, vol. 77, no. 19, pp. 10414–10422, 2003.

[77] W. E. Diehl, N. Patel, K. Halm, and W. E. Johnson, "Tracking interspecies transmission and long-term evolution of an ancient retrovirus using the genomes of modern mammals," *Elife*, vol. 5, no. MARCH2016, 2016.

[78] Á. M. Barrio *et al.*, "The first sequenced carnivore genome shows complex host-endogenous retrovirus relationships.," *PLoS One*, vol. 6, no. 5, p. e19832, Jan. 2011.

[79] B. A. Brody, S. S. Rhee, M. A. Sommerfelt, and E. Hunter, "A viral protease-mediated cleavage of the transmembrane glycoprotein of Mason-Pfizer monkey virus can be suppressed by mutations within the matrix protein," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 89, no. April, pp. 3443–3447, 1992.

[80] M. A. Sommerfelt, S. R. Petteway, G. B. Dreyer, and E. Hunter, "Effect of Retroviral Proteinase Inhibitors on Mason-Pfizer Monkey Virus Maturation and Transmembrane Glycoprotein Cleavage M13mpl9 vetraf," *J. Virol.*, vol. 66, no. 7, pp. 4220–4227, 1992.

[81] B. A. Brody, S. S. Rhee, and E. Hunter, "Postassembly Cleavage of a Retroviral Glycoprotein Cytoplasmic Domain Removes a Necessary Incorporation Signal and Activates Fusion Activity," *J. Virol.*, vol. 68, no. 7, pp. 4620–4627, 1994.

[82] M. Bobkova, J. Stitz, M. Engelstadter, K. Cichutek, and C. J. Buchholz, "Identification of R-peptides in envelope proteins of C-type retroviruses.," *J. Gen. Virol.*, vol. 83, no. Pt 9, pp. 2241–2246, Sep. 2002.

[83] M. M. Januszeski, P. M. Cannon, D. Chen, Y. Rozenberg, and W. F. Anderson, "Functional Analysis of the Cytoplasmic Tail of Moloney Murine Leukemia Virus Envelope Protein," *J. Virol.*, vol. 71, no. 5, pp. 3613–3619, 1997.

[84] S. Mi *et al.*, "Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis," *Nature*, vol. 403, p. 785, Feb. 2000.

[85] G. Cornelis *et al.*, "Retroviral envelope syncytin capture in an ancestrally diverged mammalian clade for placentation in the primitive Afrotherian tenrecs," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 41, pp. E4332–E4341, 2014.

[86] J. Denner, "Expression and function of endogenous retroviruses in the placenta," *Acta Pathol. Microbioogica Immunol. Scand.*, vol. 124, pp. 31–43, 2016.

[87] C. Lavialle *et al.*, "Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 368, no. 1626, p. 20120507, Sep. 2013.

[88] D. Lavillette, M. Marin, A. Ruggieri, F. Mallet, F. L. Cosset, and D. Kabat, "The envelope glycoprotein of human endogenous retrovirus type W uses

a divergent family of amino acid transporters/cell surface receptors," *J Virol*, vol. 76, no. 13, pp. 6442–6452, 2002.

[89] S. Blaise, N. de Parseval, L. Benit, and T. Heidmann, "Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 22, pp. 13013–13018, Oct. 2003.

[90] J. L. Blond *et al.*, "An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor.," *J. Virol.*, vol. 74, no. 7, pp. 3321–3329, Apr. 2000.

[91] C. Esnault *et al.*, "A placenta-specific receptor for the fusogenic, endogenous retrovirus-derived, human syncytin-2.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 45, pp. 17532–17537, Nov. 2008.

[92] F. Mallet *et al.*, "The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 6, pp. 1731–1736, Feb. 2004.

[93] N. de Parseval *et al.*, "Comprehensive search for intra- and inter-specific sequence polymorphisms among coding envelope genes of retroviral origin found in the human genome: genes and pseudogenes.," *BMC Genomics*, vol. 6, p. 117, Sep. 2005.

[94] A. Muir, A. M. L. Lever, and A. Moffett, "Human endogenous retrovirus-W envelope (syncytin) is expressed in both villous and extravillous trophoblast populations.," *J. Gen. Virol.*, vol. 87, no. Pt 7, pp. 2067–2071, Jul. 2006.

[95] A. Malassine *et al.*, "Expression of HERV-W Env glycoprotein (syncytin) in the extravillous trophoblast of first trimester human placenta.," *Placenta*, vol. 26, no. 7, pp. 556–562, Aug. 2005.

[96] A. Malassine *et al.*, "Expression of the fusogenic HERV-FRD Env glycoprotein (syncytin 2) in human placenta is restricted to villous cytotrophoblastic cells.," *Placenta*, vol. 28, no. 2–3, pp. 185–191, 2007.

[97] J.-L. Frendo *et al.*, "Direct involvement of HERV-W Env glycoprotein in human trophoblast cell fusion and differentiation.," *Mol. Cell. Biol.*, vol. 23, no. 10, pp. 3566–3574, May 2003.

[98] A. Vargas *et al.*, "Syncytin-2 plays an important role in the fusion of human trophoblast cells.," *J. Mol. Biol.*, vol. 392, no. 2, pp. 301–318, Sep. 2009.

[99] M. Mangeney and T. Heidmann, "Tumor cells expressing a retroviral envelope escape immune rejection in vivo.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 25, pp. 14920–14925, Dec. 1998.

[100] M. Mangeney *et al.*, "Placental syncytins: Genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 51, pp. 20534–20539, Dec. 2007.

[101] A. Dupressoir *et al.*, "Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 3, pp. 725–730, Jan. 2005.

[102] A. Dupressoir *et al.*, "Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic , endogenous retrovirus-derived ,

envelope gene," *Proc. Natl. Acad. Sci.*, vol. 106, no. 29, pp. 12127–12132, 2009.
[103] A. Dupressoir *et al.*, "A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 46, pp. E1164-73, Nov. 2011.
[104] F. Lilly, "Susceptibility to Two Strains of Friend Leukemia Virus in Mice," *Science (80-.).*, vol. 155, no. 3761, p. 461 LP-462, Jan. 1967.
[105] S. Nair and A. Rein, "Antiretroviral restriction factors in mice.," *Virus Res.*, vol. 193, pp. 130–134, Nov. 2014.
[106] T. Pincus, J. W. Hartley, and W. P. Rowe, "A major genetic locus affecting resistance to infection with murine leukemia viruses. I. Tissue culture studies of naturally occurring viruses.," *J. Exp. Med.*, vol. 133, no. 6, pp. 1219–1233, Jun. 1971.
[107] S. Best, P. Le Tissier, G. Towers, and J. P. Stoye, "Positional cloning of the mouse retrovirus restriction gene Fv1.," *Nature*, vol. 382, no. 6594, pp. 826–9, Aug. 1996.
[108] G. Boso, A. Buckler-White, and C. A. Kozak, "Ancient evolutionary origin and positive selection of the retroviral restriction factor Fv1 in muroid rodents.," *J. Virol.*, Jul. 2018.
[109] M. W. Yap, E. Colbeck, S. A. Ellis, and J. P. Stoye, "Evolution of the retroviral restriction gene Fv1: inhibition of non-MLV retroviruses.," *PLoS Pathog.*, vol. 10, no. 3, p. e1003968, Mar. 2014.
[110] P. Jolicoeur and D. Baltimore, "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.*, vol. 73, no. 7, pp. 2236–2240, Jul. 1976.
[111] P. Jolicoeur and E. Rassart, "Effect of Fv-1 gene product on synthesis of linear and supercoiled viral DNA in cells infected with murine leukemia virus.," *J. Virol.*, vol. 33, no. 1, pp. 183–195, Jan. 1980.
[112] W. K. Yang *et al.*, "Synthesis and circularization of N- and B-tropic retroviral DNA Fv-1 permissive and restrictive mouse cells.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 77, no. 5, pp. 2994–2998, May 1980.
[113] H. L. Robinson, S. M. Astrin, A. M. Senior, and F. H. Salazar, "Host Susceptibility to endogenous viruses: defective, glycoprotein-expressing proviruses interfere with infections.," *J. Virol.*, vol. 40, no. 3, pp. 745–51, Dec. 1981.
[114] J. Ito *et al.*, "Refrex-1, a Soluble Restriction Factor against Feline Endogenous and Exogenous Retroviruses," *J. Virol.*, 2013.
[115] S. Suzuki, "FV-4: a new gene affecting the splenomegaly induction by Friend leukemia virus.," *Jpn. J. Exp. Med.*, vol. 45, no. 6, pp. 473–478, Dec. 1975.
[116] J. W. Hartley, R. A. Yetter, and H. C. 3rd Morse, "A mouse gene on chromosome 5 that restricts infectivity of mink cell focus-forming recombinant murine leukemia viruses.," *J. Exp. Med.*, vol. 158, no. 1, pp. 16–24, Jul. 1983.
[117] T. Wu, Y. Yan, and C. A. Kozak, "Rmcf2, a xenotropic provirus in the Asian

mouse species Mus castaneus, blocks infection by polytropic mouse gammaretroviruses.," *J. Virol.*, vol. 79, no. 15, pp. 9677–9684, Aug. 2005.
[118] K. Kai, H. Ikeda, Y. Yuasa, S. Suzuki, and T. Odaka, "Mouse strain resistant to N-, B-, and NB-tropic murine leukemia viruses.," *J. Virol.*, vol. 20, no. 2, pp. 436–440, Nov. 1976.
[119] M. B. Gardner, C. A. Kozak, and S. J. O'Brien, "The Lake Casitas wild mouse: evolving genetic resistance to retroviral disease.," *Trends Genet.*, vol. 7, no. 1, pp. 22–7, Jan. 1991.
[120] S. Dandekar *et al.*, "Molecular characterization of the Akvr-1 restriction gene: a defective endogenous retrovirus-borne gene identical to Fv-4r.," *J. Virol.*, vol. 61, no. 2, pp. 308–314, Feb. 1987.
[121] H. Ikeda and T. Odaka, "Cellular expression of murine leukemia virus gp70-related antigen on thymocytes of uninfected mice correlates with Fv-4 gene-controlled resistance to Friend leukemia virus infection.," *Virology*, vol. 128, no. 1, pp. 127–139, Jul. 1983.
[122] H. Ikeda and T. Odaka, "A cell membrane 'gp70' associated with Fv-4 gene: immunological characterization, and tissue and strain distribution.," *Virology*, vol. 133, no. 1, pp. 65–76, Feb. 1984.
[123] C. A. Kozak, N. J. Gromet, H. Ikeda, and C. E. Buckler, "A unique sequence related to the ecotropic murine leukemia virus is associated with the Fv-4 resistance gene.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 81, no. 3, pp. 834–837, Feb. 1984.
[124] H. Ikeda, F. Laigret, M. A. Martin, and R. Repaske, "Characterization of a molecularly cloned retroviral sequence associated with Fv-4 resistance.," *J. Virol.*, vol. 55, no. 3, pp. 768–777, Sep. 1985.
[125] M. Masudal and H. Yoshikura, "Construction and Characterization of the Recombinant Moloney Murine Leukemia Viruses Bearing the Mouse Fv-4 env Gene," *J. Virol.*, vol. 64, no. 3, pp. 1033–1043, 1990.
[126] G. M. Taylor, Y. Gao, and D. A. Sanders, "Fv-4: Identification of the Defect in Env and the Mechanism of Resistance to Ecotropic Murine Leukemia Virus," *J. Virol.*, vol. 75, no. 22, pp. 11244–11248, 2001.
[127] R. S. Buller, A. Ahmed, and J. L. Portis, "Identification of two forms of an endogenous murine retroviral env gene linked to the Rmcf locus.," *J. Virol.*, vol. 61, no. 1, pp. 29–34, Jan. 1987.
[128] Y. T. Jung, M. S. Lyu, A. Buckler-White, and C. A. Kozak, "Characterization of a polytropic murine leukemia virus proviral sequence associated with the virus resistance gene Rmcf of DBA/2 mice.," *J. Virol.*, vol. 76, no. 16, pp. 8218–8224, Aug. 2002.
[129] M. S. Lyu, A. Nihrane, and C. A. Kozak, "Receptor-mediated interference mechanism responsible for resistance to polytropic leukemia viruses in Mus castaneus.," *J. Virol.*, vol. 73, no. 5, pp. 3733–3736, May 1999.
[130] Y. Anai *et al.*, "Infectious endogenous retroviruses in cats and emergence of recombinant viruses.," *J. Virol.*, vol. 86, no. 16, pp. 8634–8644, Aug. 2012.
[131] J. Ito, T. Baba, J. Kawasaki, and K. Nishigaki, "Ancestral Mutations Acquired in Refrex-1, a Restriction Factor against Feline Retroviruses,

during its Cooption and Domestication," *J. Virol.*, 2016.
[132] J. H. Blusch *et al.*, "Identification of endogenous retroviral sequences based on modular organization: proviral structure at the SSAV1 locus.," *Genomics*, vol. 43, no. 1, pp. 52–61, Jul. 1997.
[133] W. Bao, K. K. Kojima, and O. Kohany, "Repbase Update, a database of repetitive elements in eukaryotic genomes.," *Mob. DNA*, vol. 6, p. 11, 2015.
[134] D. Blanco-Melo, R. J. Gifford, and P. D. Bieniasz, "Co-option of an endogenous retrovirus envelope for host defense in hominid ancestors," *Elife*, vol. 6, pp. 1–19, 2017.
[135] O. Contributors, " 'Round-the-horn site-directed mutagenesis," *OpenWetWare*. .
[136] M. Kearse *et al.*, "Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.," *Bioinformatics*, vol. 28, no. 12, pp. 1647–1649, Jun. 2012.
[137] T. N. Petersen, S. Brunak, G. von Heijne, and H. Nielsen, "SignalP 4.0: discriminating signal peptides from transmembrane regions," *Nat. Methods*, vol. 8, p. 785, Sep. 2011.
[138] V. Blot, S. Lopez-Vergès, M. Breton, C. Pique, C. Berlioz-Torrent, and M.-P. Grange, "The conserved dileucine- and tyrosine-based motifs in MLV and MPMV envelope glycoproteins are both important to regulate a common Env intracellular trafficking.," *Retrovirology*, vol. 3, no. 1, p. 62, Sep. 2006.
[139] C. Song, S. R. Dubay, and E. Hunter, "A Tyrosine Motif in the Cytoplasmic Domain of Mason-Pfizer Monkey Virus Is Essential for the Incorporation of Glycoprotein into Virions," *J. Virol.*, vol. 77, no. 9, pp. 5192–5200, 2003.
[140] V. Cheynet, G. Oriol, and F. Mallet, "Identification of the hASCT2-binding domain of the Env ERVWE1/syncytin-1 fusogenic glycoprotein.," *Retrovirology*, vol. 3, no. 1, p. 41, Jul. 2006.
[141] B. J. Green, C. S. Lee, and J. E. J. Rasko, "Biodistribution of the RD114/mammalian type D retrovirus receptor, RDR.," *J. Gene Med.*, vol. 6, no. 3, pp. 249–259, Mar. 2004.
[142] C. S. Tailor, A. L. I. Nouri, Y. Zhao, Y. Takeuchi, and D. Kabat, "A sodium-dependent neutral-amino-acid transporter mediates infections of feline and baboon endogenous retroviruses and simian type D retroviruses.," *J. Virol.*, vol. 73, no. 5, pp. 4470–4474, 1999.
[143] S. Malicorne *et al.*, "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.*, vol. 90, no. 18, pp. 8132–8149, Sep. 2016.
[144] S. Drewlo, S. Leyting, M. Kokozidou, F. Mallet, and A. J. G. Potgens, "C-Terminal truncations of syncytin-1 (ERVWE1 envelope) that increase its fusogenicity.," *Biol. Chem.*, vol. 387, no. 8, pp. 1113–1120, Aug. 2006.
[145] V. Cheynet *et al.*, "Synthesis, assembly, and processing of the Env ERVWE1/syncytin human endogenous retroviral envelope.," *J. Virol.*, vol. 79, no. 9, pp. 5585–5593, May 2005.
[146] C. A. Kozak, "Evolution of different antiviral strategies in wild mouse

populations exposed to different gammaretroviruses," *Curr. Opin. Virol.*, vol. 3, no. 6, p. 10.1016/j.coviro.2013.08.001, Dec. 2013.
[147] N. M. Oliveira, K. B. Farrell, and M. V. Eiden, "In Vitro Characterization of a Koala Retrovirus," *J. Virol.*, vol. 80, no. 6, pp. 3104–3107, 2006.
[148] A. D. Greenwood, Y. Ishida, S. P. O'Brien, A. L. Roca, and M. V. Eiden, "Transmission, Evolution, and Endogenization: Lessons Learned from Recent Retroviral Invasions," *Microbiol. Mol. Biol. Rev.*, vol. 82, no. 1, pp. e00044-17, 2017.
[149] W. Xu *et al.*, "An exogenous retrovirus isolated from koalas with malignant neoplasias in a US zoo.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 28, pp. 11547–11552, Jul. 2013.
[150] T. Shojima *et al.*, "Identification of a novel subgroup of Koala retrovirus from Koalas in Japanese zoos.," *J. Virol.*, vol. 87, no. 17, pp. 9943–9948, Sep. 2013.