

# **The Antiretroviral APOBEC3 Proteins of Artiodactyls**

**Stefán Ragnar Jónsson**

**Ph.D. thesis**

**Supervisors:**

**Valgerður Andrésdóttir, Ph.D.**

**Reuben S. Harris, Ph.D.**



**University of Iceland**

**Faculty of Medicine**

**October 2009**



# **The Antiretroviral APOBEC3 Proteins of Artiodactyls**

Stefán Ragnar Jónsson

**Ph.D. thesis**

**Supervisors:**

**Valgerður Andrésdóttir, Ph.D.**

**Reuben S. Harris, Ph.D.**

**Doctoral Committee:**

**Ólafur S. Andrésson, Ph.D., chair**

**Eiríkur Steingrímsson, Ph.D.**

**Guðmundur Pétursson, MD**

**Faculty of Medicine**

**October 2009**

© Stefán Ragnar Jónsson  
ISBN 978-9979-9918-4-7

## ÁGRIP

APOBEC3 próteinin eru fjölskylda af-aminasa sem einungis er að finna í spendýrum. Mörg þessara próteina geta hindrað retróveirusýkingar með því að af-aminera C í U í erfðaeftni veirunnar meðan á víxlritun standur. Lentiveirur, m.a. HIV-1 hafa þróað mótsvar við þessu, veiru próteinið Vif (virion infectivity factor) sem ubiquitin merkir APOBEC3 próteinin og stuðlar að niðurbroti þeirra. Vegna vinnu okkar með *vif* gen mæði-visnuveiru vildum við athuga APOBEC3 gen og prótein sauðfjár (*Ovis aries*). Reynt var að finna varðveittar APOBEC3 raðir úr kindaerfðaeftni í gagnagrunnum. Lítið hafðist upp úr krafssinu, þannig að ákveðið var að notast einnig við efnivið úr skyldum tegundum. Þannig var hægt að finna og klóna tveggja hneppa APOBEC3 prótein úr þremur klaufdýrum, nautgripum (*Bos taurus*) og svínunum (*Sus scrofa*) sem og kindum. Tjáning þessara próteina í *E. coli* framkallaði stökkbreytingar. Þau gátu hindrað eftirmyndun HIV-1 sem og fjarskyldrar gammaretroveiru MLV, sem bendir til að bæði af-amineringarvirkni og virkni gegn retróveirusýkingum sé varðveitt. APOBEC3 prótein klaufdýra reyndust ónæm fyrir hindrun með Vif próteini HIV-1. Ólíkt hinum vel skilgreindu manna APOBEC3F og APOBEC3G þá er það N-enda af-amineringarhneppið sem er virkt í klaufdýrapróteinunum.

Raðgreining erfðamengja manna og músa hefur leitt í ljós mikinn mun í fjölda *APOBEC3* gena milli dýrategunda, frá einu í músunum til sjö í mönnum. Aðrir primatar virðast einnig hafa sjö *APOBEC3* gen. Það lá beint við að spyrja hvernig og hvenær þessi fjölgun hefur átt sér stað. Til að byrja með reyndum við að komast að fjölda *APOBEC3* gena í klaufdýrum með því að nota PCR byggt á varðveittum röðum ásamt leit í gagnabönkum en það dugði skammt. Því var brugðið á það ráð að leita að *APOBEC3* genum í BAC söfnum sem innihéldu litninga DNA kinda og svína og fullraðgreina *APOBEC3* gen þessara tegunda. Reyndust kindur hafa þrjú *APOBEC3* gen en svín tvö. Á meðan á þessu stóð var birt uppkast að erfðamengi nautgripa sem benti til að líkt og sauðfé hafi nautgripir þrjú *APOBEC3* gen. Í öllum þremur tegundum eru varðveittar leiðir til nokkurra mismunandi tjáningarforma. Þessar niðurstöður benda til þess að sameiginlegur forfaðir klaufdýra hafi haft þrjú *APOBEC3* gen og þriðja genið hafi tapast snemma í þróun *suidae* ættkvíslarinnar. Þetta bendir einnig til þess að mikill og stöðugur valþrýstingur hafi áhrif á fjölda *APOBEC3* gena mismunandi tegunda spendýra.

Komið hefur í ljós að tegundarsérhæfni leikur stórt hlutverk í samskiptum sýkils og hýsils. APOBEC3 prótein virðast veita breiðvirka vernd gegn retróveirum, en hver veira hefur þróað mótleik sem er sérsniðin að APOBEC3 próteinum viðkomandi hýsils. Tjáning á APOBEC3 próteinum úr ólíkum tegundum gæti því hugsanlega dregið úr veirusýkingu milli tegunda eða súnnum. Skortur á líffærum til líffæragjafa hefur leitt af sér notkun efniviðar úr öðrum dýrategundum til líffæragjafa í menn og hafa nokkrar slíkar aðferðir gefið góða raun. Helst er lítið til svína vegna hentugleika hvað varðar stærð, fjölda grislinga í hverju goti, stutts meðgöngutíma og hagkvæmni í ræktun. Veirusýkingar hafa þó verið talsvert til umræðu eftir að sýnt var fram á að innrænar retróveirur úr svínunum (PERV) geta sýkt mannafrumur í frumurækt. Aðferðir til að draga úr hættu á sýkingum eru því eftirsóknarverðar. Við vildum láta reyna á hvort tjáning á APOBEC3 próteini úr annarri dýrategund gæta dregið úr eða hindrað færslu veira frá svínunum til manna. Samræktunartilraunir voru notaðar til að sýna fram á að tjáning manna APOBEC3 í veiru-framleiðandi

svínafrumum kom í veg fyrir sýkingu yfir í mannafrumur. APOBEC3G hindraði PERV áður en veiran var innlimuð í erfðaeefni mannafrumanna. Hins vegar virtist APOBEC3 prótein svína ekki geta komið í veg fyrir sýkingu svínaveirunnar. Þetta bendir til að nota mætti tjáningu próteina milli tegunda til framleiðslu á hættuminni efnivið fyrir líffæragjafir.

Lykilorð: Lentiveira, mæði-visnu veira, APOBEC3, cýtósín afaminering, retróveiru hindrun

## ABSTRACT

The APOBEC3 proteins are a mammalian specific family of nucleic acid cytosine deaminases. A number of family members, including the most studied family member human APOBEC3G, have been shown to be able to inhibit retroviral infection by causing C to U deamination of retroviral DNA during reverse transcription. Some lentiviruses, including HIV-1, can counteract this defense through the viral protein Vif (viron infectivity factor) that triggers APOBEC3 protein ubiquitination and degradation. Because of our prior work with maedi-visna virus our goal was to identify the APOBEC3 genes and proteins of sheep (*Ovis aries*). This was done by searching for sheep material in databases using conserved motifs from known APOBEC3 proteins. This was challenging since there was very little sheep sequence available and APOBEC3 proteins are under strong positive selection which makes for less conservation of sequence. By utilizing available sequences from related artiodactyl species, we could identify and characterize double domain APOBEC3 proteins from three artiodactyls (cloven hoof ungulates), cattle (*Bos taurus*), pig (*Sus scrofa*) as well as sheep. Expression of all of these proteins in *E. coli* triggered a mutator phenotype. They were also capable of blocking the replication of HIV-1 and to a lesser extent, MLV a distantly related gammaretrovirus of mice, indicating that both the deaminase and the anti-retroviral activities were conserved. The dinucleotide preferences of these proteins were broader than the highly specific preferences of the well studied human APOBEC3F and APOBEC3G. The artiodactyl double-domain APOBEC3 proteins were resistant to inhibition by HIV-1 Vif. Surprisingly, unlike the well characterized human proteins, the deamination activity of these 3 artiodactyl double domain proteins resides in the N-terminal deamination domain.

The genomic sequences of mice and humans revealed a dramatic expansion in the number of *APOBEC3* genes – from one in mice to seven in humans. Other primates also appear to have seven *APOBEC3* genes. An obvious question was therefore whether this expansion occurred gradually during the evolution of mammals or whether it occurred recently. To begin to address this question, we used degenerate PCR, RACE and database searching to assess the number of *APOBEC3* genes in the artiodactyl species we had worked with cattle, sheep and pigs. Several *APOBEC3* sequences were recovered but these approaches proved far from definitive. We therefore used APOBEC3-specific hybridization to identify pig and sheep BACs harboring this locus. The full DNA sequence spanning this genomic region was determined by sequencing BAC fragments. Interestingly, sheep proved to have three *APOBEC3* genes, whereas pigs had only two. During the course of this work, the cattle genome project released a draft sequence, which also indicated the presence of three *APOBEC3* genes. In all three species, read-through transcription and alternative splicing also produced a catalytically active double domain protein, bolstering the innate antiretroviral defenses. These data therefore suggest that the common ancestor of primates and artiodactyls had three *APOBEC3* genes and that during mammalian evolution the *suidae* lineage lost one. These data further suggest the presence of an ongoing selective pressure that has facilitated the gradual expansion of *APOBEC3* gene number during the evolution of mammals.

An emerging theme in APOBEC3 research is species specificity. APOBEC3 proteins seem to have a broad inhibitory effect of retroviruses, but each retrovirus seems to have evolved counter-measures that are specific to the APOBEC3 proteins of their host. The usage of cross-species expression might therefore be beneficial in raising barriers to impede cross species viral transmissions or zoonosis. The shortage of allogenic material for transplantation has led to the possible use of materials of non-human origin for transplantation purposes (Xenotransplantation). An unprecedented opportunity for zoonosis occurs when live cells, tissues or organs are transplanted from one species to another. Several xenotransplantation procedures have shown preclinical promise for treating a number of human conditions. Pigs are favorable xenotransplantation sources because of their human-like physiology, large litters, short gestation period and genetic malleability. However, pig to human virus transmission has been a concern since it was shown that porcine endogenous retroviruses (PERVs) could infect human cells in culture. Methods that reduce PERV transmission and possible xenozoonotic infections are therefore desirable. We conceived that introducing a foreign APOBEC3 protein might impede the transmission of viruses from pigs to humans. Long-term co-culture experiments were used to show that PERV transmission from pig to human cells is reduced to nearly undetectable levels by expressing human APOBEC3G in virus-producing pig kidney cells. APOBEC3G interferes with PERV replication before the virus immortalizes by integrating into human genomic DNA. In contrast the endogenous APOBEC3 proteins of pigs appeared insufficient to inhibit PERV. These studies indicate that cross-species restriction strategies may help provide safer xenotransplantation resources.

Keywords: Lentivirus, maedi-visna virus, APOBEC3, cytosine deamination, retrovirus restriction

## ACKNOWLEDGMENTS

The research described in this thesis was performed at the Institute for Experimental Pathology at the University of Iceland and at the Department of Biochemistry, Molecular Biology and Biophysics at the University of Minnesota.

I would like to thank my supervisors, Valgerður Andrésdóttir for allowing me the chance to work on this project as well as her excellent supervision and support and the courage to send me overseas and allowing me to stay there for much longer than originally planned and Reuben Harris, for inviting me into his then recently started lab, his supervision and encouragement and for approaching artiodactyls with an open mind.

The members of my doctoral committee: Ólafur S. Andrésen, Eiríkur Steingrímsson and Guðmundur Pétursson for their advice during my studies.

I would like to thank the staff at Keldur for the co-work through the years. Special thanks to members of the “visna group” for their help and insightful comments at lab meetings, and thanks to my fellow students for making the lab a fun place at all times.

I would like to thank the members of the Harris lab especially Rebecca, Guylaine, Mark, April, Donna and Bill for making the Harris lab a great place to work. I would also like to thank members of neighboring labs for making the Minnesota experience great and of course for the ever joyful Friday lab meetings.

And last but certainly not least I would like to thank my family for their everlasting support and belief in me and my work.

This work was funded by the Icelandic Research Council, the Icelandic Fund for Graduate Students, the Research Fund of the University of Iceland, the University of Minnesota Val Bjornson Icelandic Exchange Scholarship, the University of Minnesota, the NIH and Spring Point Project.

## CLARIFICATION OF CONTRIBUTION

**Paper I** – This paper describes the delineation of the molecular basis for the differential growth phenotypes of two molecular clones of maedi-visna virus. Part of this study was the MS project of Bjarki Guðmundsson at the Faculty of Life- and Environmental Sciences at the University of Iceland. I participated in virus isolations from sheep blood and organs and primary sheep macrophage isolations and infections. I designed the real-time PCR assay used for replication kinetics studies. I read and made comments on the manuscript.

**Paper II** – This paper describes the cloning and characterization of double-domain APOBEC3 proteins from cattle, sheep and pigs. I cloned the APOBEC3 proteins and performed the phylogenetic analysis. Performed the *E. coli* based mutation assays and the GFP-based retrovirus infection assays. I collected and analyzed the data. Guylaine Haché constructed the catalytic mutants and Mark D. Stenglein performed the fluorescent microscopy. I participated in writing the manuscript.

**Paper III** – This paper describes the sequencing of the full *APOBEC3* genetic loci of pigs and sheep as well as the coding potential of the sheep, cow and pig *APOBEC3* genes. The work on which this paper is based on was led by Rebecca S. LaRue. I participated in collecting and editing the sequencing data on the sheep *APOBEC3* locus and performing the *E. coli* based assay for deamination activity. I read and made comments on the manuscript at multiple stages.

**Paper IV** – This paper describes the inhibition of PERV transfer from pig to human cells by expressing human APOBEC3G in the virus producing cells. I designed the long-term co-culture system and generated the stable cell lines. Performed the co-culture assays and collected the DNA for PCR and sequencing. I analyzed the data. Rebecca S. LaRue performed the qPCR studies and the catalytic mutant co-culture and Mark D. Stenglein performed the fluorescent microscopy. I participated in writing the manuscript.

# CONTENTS

<i>ÁGRIP</i>	<i>i</i>
<i>ABSTRACT</i>	<i>iii</i>
<i>ACKNOWLEDGMENTS</i>	<i>v</i>
<i>CLARIFICATION OF CONTRIBUTION</i>	<i>vi</i>
<i>CONTENTS</i>	<i>vii</i>
<i>LIST OF FIGURES AND TABLES</i>	<i>x</i>
<i>LIST OF PAPERS</i>	<i>xi</i>
<i>LIST OF ABBREVIATIONS</i>	<i>xii</i>
<b>1. INTRODUCTION</b>	<b>1</b>
History of maedi-visna virus	1
Strains of MVV	3
Retroviruses	5
Genome and proteins	7
Gag	8
Pol	9
Env	10
Additional virus proteins	10
Rev	10
Tat	10
Vif	11
Non MVV auxiliary proteins	13
Receptor binding, entry and uncoating	13
Reverse transcription	14
Nuclear entry and integration	16
Transcription	16
Expression of viral proteins and virion formation	17
Viron assembly	19
Maturation	20
Endogenous retroelements [and retroviruses]	20
Porcine endogenous retroviruses (PERVs) and xenotransplantation	24
APOBEC3 protein family of cytosine deaminases	28
The APOBEC3 family cytosine deaminases	28
Inhibition of exogenous and endogenous retroelements	32
Deamination-independent inhibition	33

Modes of viral escape	35
APOBEC3 evolution and positive selection	36
<b>3. MATERIALS AND METHODS</b>	<b>39</b>
Cells and cell-culture	40
Macrophage isolation and culture	40
Virus and cells	40
Generation of stable cell lines	40
Long-term co-culture experiments	41
DNA methods	42
Artiodactyl A3F cDNA sequences and expression plasmids	42
A3Z1 gene degenerate PCR analyses	44
Real-time PCR assay for MVV	44
Genomic DNA Quantitative-PCR (qPCR) assays	45
PERV <i>pol</i> Gene DNA sequence analyses	45
RT-PCR experiments	46
APOBEC3 loci genomic DNA sequences	46
Amino acid alignments and phylogenetic analyses	47
Infections and Infectivity assays	47
Experimental infection of sheep.	47
Virus isolation from blood and organs	47
Green fluorescent protein (GFP)-based retrovirus infection assays	47
Retroviral DNA sequence analyses	48
Reverse Transcriptase assay	48
<i>E. coli</i> -based deamination assay	49
Fluorescent Microscopy	50
<b>4. RESULTS</b>	<b>51</b>
Simultaneous mutations in CA and Vif of Maedi-visna virus cause attenuated replication in macrophages and reduced infectivity in vivo (paper I)	51
The growth attenuation of the KS1 molecular clone is determined by regions of Vif and CA.	51
In vivo inoculation with the two parental strains and the recombinant virus CA(L120R)-Vif(P205S).	51
Evolutionary conserved and non-conserved retrovirus restriction activities of artiodactyl double-domain APOBEC3 (APOBEC3F) proteins (paper II)	54
The double deaminase domain APOBEC3 proteins of artiodactyls	55
The artiodactyl double domain A3 proteins catalyze DNA cytosine deamination	57
The artiodactyl double domain A3 proteins are predominantly cytoplasmic	58
Retrovirus restriction by double domain artiodactyl A3 proteins	59

The N-terminal zinc-binding, deaminase domain of the artiodactyl double domain A3 proteins catalyzes C → U deamination, and this activity is necessary for full levels of retrovirus restriction	61
Retroviral hypermutation properties of artiodactyl double domain A3 proteins	62
The artiodactyl double domain A3 proteins are fully resistant to HIV-1 Vif	63
The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals (paper III)	64
Sheep and cattle have three A3 genes with a Z1-Z2-Z3 organization	64
The pig has two A3 genes with a Z2-Z3 organization	65
The restriction of Zoonotic PERV transmission by human APOBEC3G (paper IV)	67
A Co-culture Assay to Monitor PERV Transmission	67
Human APOBEC3G Inhibits PERV Transmission	68
The Sub-cellular Distribution of Human APOBEC3G Is Virtually Identical in Human and Pig Cell Lines	70
PERV Appears Resistant to Porcine APOBEC3Z2-Z3	71
Evidence that Human APOBEC3G Inhibits PERV By a Deamination-Independent Mechanism	71
Genetic Variation in Zoonosed PERV DNA Sequences	72
<b>5. DISCUSSION</b>	<b>74</b>
<b>6. REFERENCES</b>	<b>80</b>
<b>7. PAPERS I – IV</b>	<b>96</b>

## LIST OF FIGURES AND TABLES

<b>Figure</b>	<b>Title</b>	<b>Page</b>
Figure 1	Strains of maed-visna virus.....	5
Figure 2	Phylogeny of retroviruses.....	7
Figure 3	Genomes of lentiviruses.....	8
Figure 4	Reverse transcription.....	15
Figure 5	Lentiviral replication.....	18
Figure 6	APOBEC3 mediated inhibition of lentiviral infection.....	33
Figure 7	Schematics of the APOBEC3 repertoires of mammals whose genomes have been sequenced.....	37
Figure 8	Co-culture system for detection of PERV transmission.....	41
Figure 9	Diagrammatic representation of recombinant MVV derived from the molecular clones KV1772 and KS1.....	52
Figure 10	Growth of viruses in macrophages from 4 different sheep.....	53
Figure 11	A comparison of artiodactyl A3F proteins.....	56
Figure 12	DNA cytosine deaminase activity of the artiodactyl A3Z2-Z3 proteins in <i>E.coli</i> .....	58
Figure 13	Sub-cellular distribution of the artiodactyl A3Z2-Z3 proteins in comparison to the orthologous human and mouse A3 proteins....	59
Figure 14	Retrovirus restriction activity of the artiodactyl A3Z2-Z3 proteins...	60
Figure 15	Relative contributions of the N- and C-terminal zinc-binding domains to HIV-GFP restriction.....	63
Figure 16	Artiodactyl A3Z2-Z3 proteins are resistant to HIV-1 Vif.....	64
Figure 17	The <i>APOBEC3</i> genomic repertoire of sheep, cattle and pigs.....	66
Figure 18	PERV Transmission Assay.....	68
Figure 19	Human APOBEC3G Inhibits PERV Transmission.....	69
Figure 20	The Sub-cellular Distribution of Human APOBEC3G in Human and Pig Cell Lines.....	70
Figure 21	The APOBEC3G-dependent Inhibition of PERV Transmission Is Deamination-Independent.....	72
Figure 22	Genetic Variation in Zoonosed PERV DNA Sequences.....	73
<b>Table</b>	<b>Title</b>	<b>Page</b>
Table 1	Summary of restriction patterns by APOBEC3 proteins.....	34
Table 2	Summary of restriction patterns and subcellular localization of APOBEC3 proteins.....	36
Table 3	Frequency of virus isolation from KV1772-, KS1-, and CA(L120R)-Vif(P205S)-infected sheep.....	54

## LIST OF PAPERS

The thesis is based on the following papers

### Paper I

Bjarki Gudmundsson, Stefán Ragnar Jónsson, Oddur Ólafsson, Guðrún Agnarsdóttir, Sigríður Matthíasdóttir, Gudmundur Georgsson, Sigurbjörg Torsteinsdóttir, Vilhjálmur Svansson, Helga Bryndís Kristbjörnsdóttir, Sigríður Rut Franzdóttir, Ólafur S. Andrésson, and Valgerður Andrésdóttir. 2005. Simultaneous Mutations in CA and Vif of Maedi-Visna Virus Cause Attenuated Replication in Macrophages and Reduced Infectivity in Vivo. *Journal of Virology* 79: 15038-15042

### Paper II

Stefán R. Jónsson, Guylaine Haché, Mark D. Stenglein, Scott C. Fahrenkrug, Valgerður Andrésdóttir and Reuben S. Harris. 2006. Evolutionary conserved and non-conserved retrovirus restriction activities of artiodactyl APOBEC3F proteins. *Nucleic Acids Research*, Vol. 34 No. 19 pp. 5683-5694

### Paper III

Rebecca S. LaRue, Stefán R. Jónsson, Kevin. A.T. Silverstein, Mathieu Lajoie, Denis Bertrand, Nadia El-Mabrouk, Isidro Hotzel, Valgerður Andresdottir, Timothy P.L. Smith and Reuben S. Harris. 2008. The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. *BMC Molecular Biology*, 9:104  
doi:10.1186/1471-2199-9-104

### Paper IV

Stefán R. Jónsson, Rebecca S. LaRue, Mark D. Stenglein, Scott C. Fahrenkrug, Valgerður Andrésdóttir and Reuben S. Harris. 2007. The Restriction of Zoonotic PERV Transmission by Human APOBEC3G. *PLoS ONE*. 2(9): e893.  
doi:10.1371/journal.pone.000893

Reprints have been made with the permission of the publishers

## LIST OF ABBREVIATIONS

APOBEC	Apolipoprotein B mRNA editing enzyme catalytic polypeptide
BAC	Bacterial Artificial Chromosome
BIV	Bovine immunodeficiency virus
BLAST	Basic Local Alignment Search Tool
bp	base pairs
CA	Capsid protein
CAEV	Caprine Arthritis Encephalitis virus
CMV	Cytomegalovirus
DMEM	Dulbecco's modification of Eagles' medium
DNA	Deoxyribonucleic acid
eGFP	Enhanced green fluorescent protein
EIAV	Equine infectious anemia virus
Env	Envelope protein
EST	Expressed Sequence Tag
FACS	Fluorescence-activated cell sorting
FIV	Feline immunodeficiency virus
FOS	Fetal ovine synovial cells
HERV	Human endogenous retrovirus
HIV	Human immunodeficiency virus
HTLV-1	Human T-cell lymphotropic virus
IN	Integrase
JDV	Jembrana disease virus
JSRV	Jaagsiekte sheep retrovirus
kb	Kilo base
KoRV	Koala retrovirus
LTR	Long terminal repeats
MA	Matrix
MLV	Murine leukemia virus
mRNA	Messenger RNA
MVV	Maedi-visna virus
MYA	Million Years ago
NC	Nucleocapsid protein
nt	Nucleotide
PBS	Primer binding site
PCR	Polymerase chain reaction
PERV	Porcine endogenous retrovirus
PIC	Preintegration complex
Pol	Polymerase
PR	Protease
qPCR	Quantitative polymerase chain reaction
RELK	Rabbit endogenous lentivirus type K
RNA	Ribonucleic acid
RRE	REV responsive element
RT	Reverse transcriptase
SIV	Simian immunodeficiency virus
SU	Surface protein
TAR	Tat activation region
Tat	Trans-activation protein

TM	Trans membrane protein
UTR	Untranslated region
Vif	Virion infectivity factor
VLP	Viral like particle
Vpr	Viral protein R
Vpu	Viral protein U
VSV-G	Vesicular stomatitis virus protein G

#### Abbreviations of species names

Bt	Bos taurus	cow
Cf	Canis familiaris	dog
Fc	Felis catus	cat
Mm	Mus musculus	mouse
Oa	Ovis aries	sheep
Pt	Pan troglodytes	chimpanzee
Rn	Rattus norvegicus	rat
Ss	Sus scrofa	pig

## 1. INTRODUCTION

### **History of maedi-visna virus**

In the mid 1930s a previously unknown disease became apparent in Icelandic sheep in two different regions of the country (Coffin, 1992; Pálsson, 1976).

This new disease manifested itself as dyspnoea (laboured breathing) and was always fatal. The disease, known for its long incubation period and insidious onset, mainly occurred in adult sheep older than 2-3 years old and started with the animals being dyspnoeic under exposure to strenuous activity. The disease was called maedi (mæði, Icelandic for “shortness of breath” or “laboured breathing”) and also “dry maedi” to differentiate it from another lung disease in sheep which was called “wet maedi” (which turned out to be pulmonary adenomatosis caused by the jaagsiekte sheep retrovirus (JSRV)), which had been observed a few years earlier in one of the regions affected by this new disease. Both of these diseases were brought to Iceland by a flock of sheep of the Karakul breed imported from Halle, Germany in 1933. The sheep appeared healthy at inspection and showed no signs of any diseases during a period of two months in quarantine. It is interesting to note that sheep of the Karakul breed in Germany never developed any signs of disease corresponding to the ones they brought into the Icelandic sheep population (Straub, 2004). This flock of German sheep brought yet another disease that was only seen in sheep in the southwest region of the country which was one of the two regions where maedi was seen as well (the other was the north-eastern part). This was a neurological disease which also had a long incubation period and was very insidious in its onset. It caused encephalomyelitis with weakness in the hind limbs leading to paraplegia and even to complete paralysis. It was named “visna” (Icelandic for wasting). It was rarely seen and only came up in flocks that were also infected by maedi. In rare cases visna was the more common disease (Pálsson 1976).

In 1937 an eradication scheme was initialized to rid Iceland of maedi. The country was divided into quarantine zones, and all the sheep in the infected zones were culled, the farms left without sheep for three years and then restocked with sheep from unaffected quarantine zones. It is estimated that up to 650.000 sheep were culled during this process and approximately 150.000 sheep died from the disease itself (Pálsson 1976). In the years from 1937-1944 the Icelandic sheep

population went from 730.000 to 450.000 winterfed sheep (Pálsson 1976). The culling scheme was successful and the last infected flock was put down in 1965. To this day Iceland is the only nation in the world that has successfully eradicated maedi-visna virus or any lentiviral epidemic.

To battle these emerging diseases of the primary farming animal, the government set-up (with generous funds from the Rockefeller Institute) an Institute of Experimental Pathology affiliated with the medical department of the University of Iceland in 1948. It was situated at a farm that the government had purchased in the outskirts of the capital, Reykjavík. The Institute was known by the name of the farm which had stood on the land, Keldur. The first director of the Institute was a young Icelandic physician, Dr. Björn Sigurðsson. His work and that of his colleagues laid the foundation for the scientific understanding of the new diseases. By serial passages of homogenates from brain and lungs they found out that the diseases were caused by a filterable agent, most probably a virus (Sigurdsson et al., 1957; Sigurdsson et al., 1953). A hallmark of the experimental infection experiments was the long incubation period. In most cases several months passed until any symptoms of the disease were seen in the animals. Upon examination of organs at earlier time points, microscopic changes in the affected organs were seen as well as an increase of cells in the cerebral spinal fluid. In tissue cultures however the agents caused syncytia formation and cell death in a matter of a few days (Sigurdsson et al., 1960) .

Based on Sigurðsson's experience with these diseases, maedi and visna, as well as rida (scrapie), he found that they did not fit into the accepted classification of diseases at that time, either acute or chronic, due to their long incubation times and very classical disease path that always led to a serious disease or death even though the host immune system should have had ample time to neutralize the disease agent. He therefore set forth a new class of disease and named them slow infections based on the following identifications (Sigurdsson, 1954a; Sigurdsson, 1954b):

1. A very long period of latency, lasting from several months to several years.
2. Once clinical symptoms have appeared they last a long time, ending, usually in serious disease or death.
3. Each of the atypically slow diseases is limited to a single host species and pathological changes are usually confined to a single organ or tissue system.

Into this new class of diseases he suggested that along with maedi and visna could be classified diseases known at that time, like infectious adenomatosis of sheep lungs (caused by jaagsiekte sheep retrovirus, JSRV a betaretrovirus), rida or scrapie (which was originally thought to be caused by a virus but was later found out to be caused by apparently folded prion proteins (PrP<sup>c</sup> → PrP<sup>sc</sup>) (Prusiner, 1982; Prusiner and Scott, 1997), cancer of the udder in mice (caused by mouse mammary tumour virus, MMTV a betaretrovirus) and Rous sarcoma in chickens (caused by Rous sarcoma virus, an alpharetrovirus and one of the prototype retroviruses, first characterized by Peyton Rous in 1911 (Rous, 1911; Sigurdsson and Palsson, 1958)).

Diseases corresponding to those by MVV have been described in a number of countries, with pneumonia being the most prevalent form. Countries affected by maedi-like diseases include the USA, most countries in Europe, South Africa, Ethiopia, Kenya and Kirgizstan (Thormar, 2005).

### **Strains of MVV**

MVV strains that have been isolated can be divided roughly into clinical and lab strains where the clinical strains are isolates from natural infections and the lab strains isolated from sheep that have been experimentally infected or prepared by genetic manipulations. Some of these strains cause maedi only and some cause both maedi and visna. Strains isolated from naturally infected sheep are diverse since the virus is known in geographically distinct locations and has probably resided in the sheep population for a very long time.

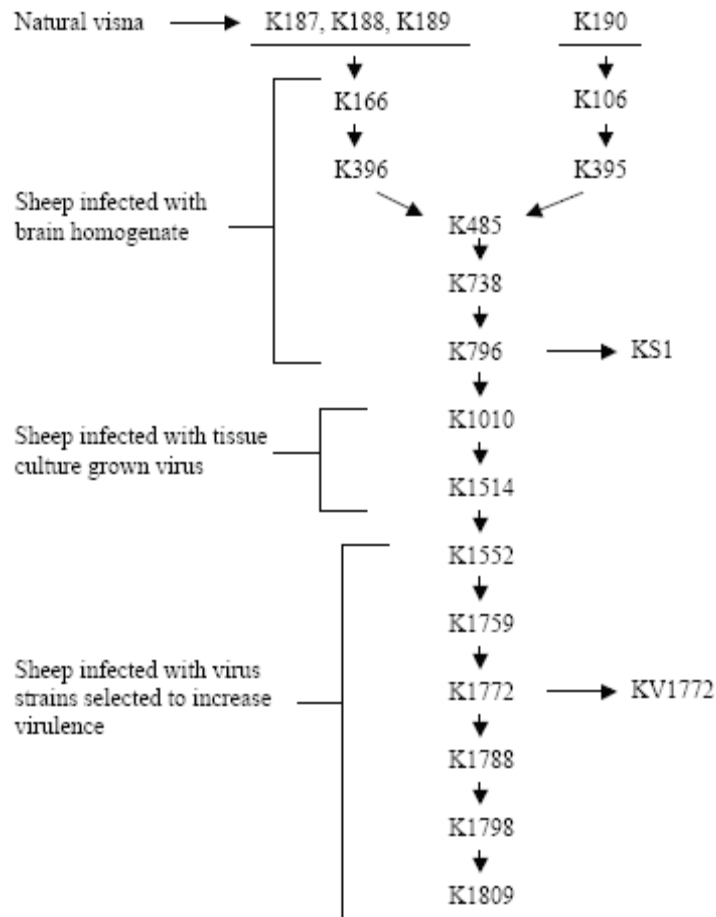
In scientific work the most widely used strains are visna-causing strains that have been propagated both *in vivo* and in cell culture for many years. The origins of many of them date back to the initial experimental infections that originated in 1949 where highly infectious and neurotropic strains were isolated (Figure 1). Examples of such strains are K1514, which is widely used, and KV1772, which is a derivative of K1514 and was selected for increased infectivity and neuroinvasiveness. An infectious molecular clone derived from this strain yields highly infectious virus *in vivo* and in cell culture (Andresson et al., 1993). The strain LV1-KS1 (commonly known as KS1) is also a molecular clone and was isolated in similar fashion as KV1772 (Staskus et al., 1991). Although genetically very similar to KV1772, KS1 is very different when it comes to infectivity, both *in vitro* and *in vivo*, when infecting

Icelandic sheep (Torsteinsdottir et al., 1997). The two molecular clones differ by 129 nucleotide substitutions and two deletions of 3 and 15 nucleotides in the *env* gene. KS1 replicates well in SCP (sheep choroid plexus) cells, but replicates only to low titers in macrophages, and is non-pathogenic in sheep. KV1772 replicates to high titers in SCP cells and macrophages and is highly pathogenic in sheep. By constructing recombinant viruses and a series of amino acid substitution mutations, the attenuated phenotype of KS1 was mapped to two simultaneous point mutations (as neither one alone was sufficient to convert KV1772 to a growth restrictive phenotype), one in the CA domain of *gag* and the other in *vif* (Gudmundsson et al., 2005, paper I). The ability of KS1 as an attenuated vaccine to induce protection after intratracheal infection against superinfection with the highly pathogenic KV1772 was tested by infecting 4 sheep with KS1. Ten months later these sheep were challenged with KV1772 and another 4 unvaccinated sheep of the same age were infected simultaneously. All sheep became infected with the challenge virus. The intratracheal vaccination with KS1 did not protect against the challenge superinfection. However, virus was isolated more frequently from the blood from the unvaccinated controls than of the vaccinated animals and ten times more frequently from lungs of unvaccinated sheep than from lungs of vaccinated sheep at sacrifice, indicating some level protection (Petursson et al., 2005).

Maedi was the most prevalent disease form in the maedi-visna epizootic in Iceland from 1933 to 1965. However, in some flocks visna was the main cause of disease and death. Studies have shown that the neurovirulent strains have a duplication in the U3 region of the MVV LTR which is a determinant of cell tropism (Agnarsdottir et al., 2000). A deletion of a CAAAT sequence from either one of the repeats resulted in poor virus growth in sheep choroid plexus (SCP) cells. A duplication in the LTR encompassing the CAAAT sequence was found in four neurological field cases that were sequenced, but no duplication was present in the LTRs from seven maedi cases, indicating that the duplication in the LTR is associated with neurovirulence (Oskarsson et al., 2007).

Many MVV strains have been fully sequenced. In 1985 Gonda and colleagues identified a remarkable sequence and morphological similarity between MVV and what was called HTLV-III (later known as HIV-1). The similarities were greater than to other known viruses, and it was therefore suggested that HIV-1 was related to the lentiviruses (Gonda et al., 1985). The sequencing of the K1514 genome

further demonstrated the relationship between the viruses, especially based on genome organization and sequence homologies in the *pol* gene (Sonigo et al., 1985). Based on these similarities HIV-1 was classified as a lentivirus. Since then other lab strains of MVV have been fully sequenced such as the widely used KS1 and KV1772 among others (Andresson et al., 1993; Staskus et al., 1991).



**Figure 1.** Strains of maedi-visna virus. Lineage of maedi-visna virus strains isolated for increased infectivity and pathogenicity.

### Retroviruses

The maedi-visna virus belongs to the *retroviridae* family. Retroviruses are a family of single strand RNA viruses that reverse transcribe their genetic material into double stranded DNA and integrate it into the genome of the host cell. The viral genome is a single strand of (+) RNA 7-12 kb in length. All retroviruses have two linked copies of gRNA in the mature virion (Lin and Thormar, 1970; Vogt, 1997)

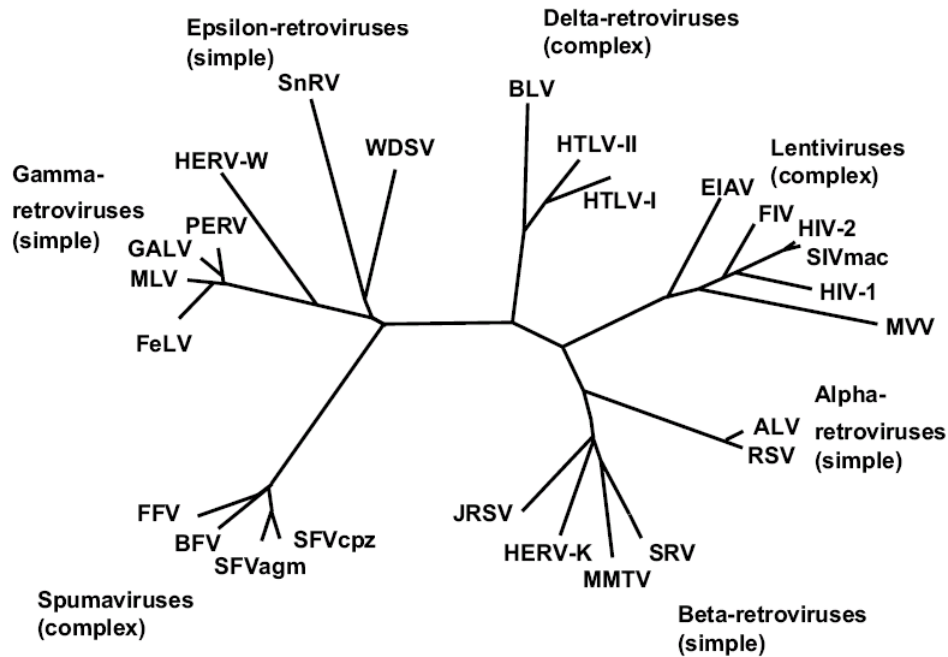
The *retroviridae* is divided into three main groups, *oncoviridae* or oncoviruses (i), *spumaviridae* or spuma- or foamyviruses (ii) and the *lentiviridae* or lentiviruses (iii) (Figure 2).

*Oncoviridae* or oncoretroviruses are tumour causing viruses divided into five subdivisions, *alpha*  $\alpha$  (simple tumour causing viruses, very host-specific, mainly found in birds. Examples of alpharetroviruses are RSV and ALV), *beta*  $\beta$  (simple oncogenic retroviruses that cause tumours in a variety of species, examples are Mouse Mammary Tumour virus, MMTV, Jaagsiekte Sheep Retrovirus, JSRV, and Mason-Pfizer Monkey virus), *delta*  $\delta$  (complex retroviruses that cause leukaemias in a number of species. The best known are, Human T cell lymphotropic virus type 1, HTLV-1 and Bovine Leukaemia virus, BLV); *gamma*  $\gamma$ , the largest oncovirus subgroup containing a large number of both endo- and exogenic viruses in a large variety of species. The best known example is the Murine Leukaemia virus MLV. Other gamma retroviruses include Gibbon ape leukaemia virus GALV in gibbons, Feline leukaemia virus FeLV in cats and porcine endogenous retrovirus PERV in pigs, and *epsilon*  $\epsilon$  (the smallest group of oncoviruses and the least studied, found mainly in fish. The best known are WDSV, walleye dermal sarcoma virus and SnRV or snakehead retrovirus).

*Spumaviridae* or spuma- or foamyviruses are widespread viruses (the best studied one being a chimpanzee derived virus termed PFV or primate foamy virus but others include HFV human foamy virus, BFV bovine foamy virus and FFV feline foamy virus) not known to cause disease and are in many ways unconventional retroviruses when it comes to both infectivity and replication (they do not process or cleave their Gag protein into smaller sub-units, the NC part of the foamy virus Gag protein does not contain any zinc fingers and reverse transcription is nearly completed before the virus buds from a producer cell).

*Lentiviridae* or lentiviruses are the most complex of the retroviruses. Lentiviruses are found in a number of mammalian species; HIV-1 and HIV-2 (human immunodeficiency virus) in humans, SIV (simian immunodeficiency virus) found in a wide variety of simian species (virus strains/subtypes have been isolated from around 30 species), BIV (Bovine immunodeficiency virus) in cows, MVV (maedi-visna virus) in sheep, CAEV (caprine arthritis encephalitis virus) in goats, EIAV (equine infectious anaemia virus) in horses and FIV (feline immunodeficiency virus) in cats. The name lentivirus was based on Dr. Sigurdsson's concept of slow

infections (lentus meaning slow in Latin) (Haase, 1975). They do not cause cell transformations but are the cause of diseases with a long incubation period usually leading to death. Retroviruses have been extensively studied for decades because of their involvement in tumour formation and transformation of cell lines, and in the last 25 years even more so due to the AIDS pandemic. HIV-1 has become the most extensively studied 9000 nucleotides in history. Regardless of intense study, HIV continues to be a major modern plague threatening humankind.



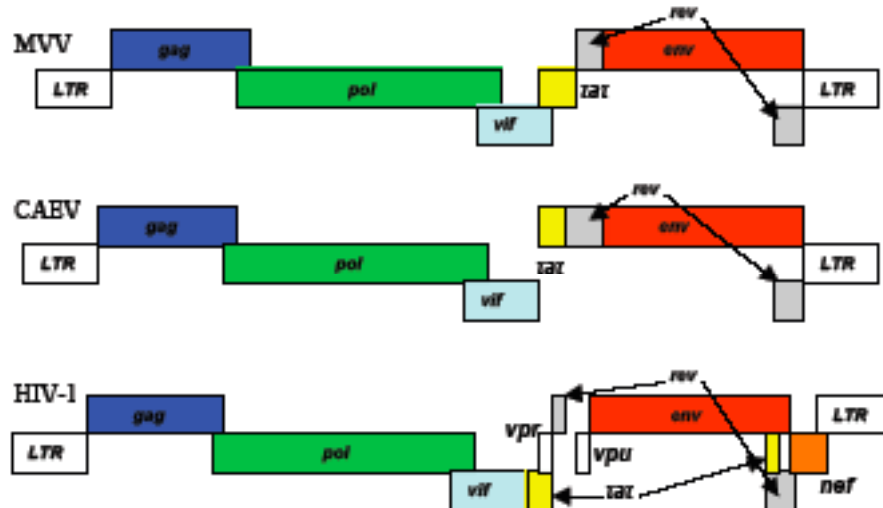
**Figure 2.** Phylogeny of retroviruses. An un-rooted phylogenetic tree of retroviruses based on homology of pol sequences (from Weiss et al 2006).

### Genome and proteins

The overall genome organization of retroviruses is conserved. All encode the major building blocks of the virus particle, Gag (group specific antigen), Pol (polymerase) and Env (envelope). Along with Gag, Pol and Env, lentiviruses also encode Tat and Rev. All but EIAV have Vif. Non-primate lentiviruses have dUTPase as part of the *pol* gene, and the primate lentiviruses have Vpr, Nef. HIV-1 and SIV<sub>cpz</sub> further have Vpu, whereas HIV-2 and most SIV viruses have Vpx.

The LTR sequences on either end of the genome are important for integration, reverse transcription and transcription regulation. The repeats are divided into three regions, U3, R and U5. U3 (unique) is at the 3' end of the RNA and U5 is at the 5' end but R (repeat) is on both ends of the RNA genome. During reverse transcription two copies of the repeats are formed in two jumps, one on each end of

the genome. The primer binding site where the tRNA primer binds is just outside the U5 of the LTR. The U3 of the LTR contains the TATA box and other promoter and enhancer sequences. The polyadenylation site for mRNA is between the R and U5 at the 3' end (Sonigo et al., 1985; Vigne et al., 1987; Vigne et al., 1994; Vogt, 1997). The MVV genome is 9.2 kb (Sonigo et al., 1985) (Figure 3).



**Figure 3.** Genomes of lentiviruses. Comparison of the genomes of MVV, CAEV and HIV-1. Based on (Clements and Zink, 1996)

### Gag

The *gag* gene encodes a 55 kDa protein, Gag<sup>P55</sup> or P55Gag, a polyprotein precursor that makes up the building blocks of the virion and regulates the assembly and formation of virions, 2000-4000 Gag copies are found in each virion (Coffin, 1992). Gag is further divided into three subunits, from the N end: matrix (MA, p17), capsid (CA, p25) and nucleocapsid (NC, p14) (Pepin et al 1998). The viral Protease cleaves Gag into its mature subunits.

The matrix protein MA controls Gag assembly and attachment to the cell membrane. The membrane affinity is governed by N-end myristylation (Pepin et al 1998). The protein forms trimers that attach to the inside of the membrane. MA is part of the RT- and preintegration complex (Vogt 1997).

CA is 200-270 amino acids, and it is rich in helices. CA contains the major homology region (MHR) that is conserved among all retroviruses, with the exception of spumaviruses. CA forms a tight hydrophobic core that encloses the virus' genetic material (Freed, 2001; Vogt, 1997).

NC is small basic protein, 79 amino acids in MVV (Andresson et al., 1993), that packages the viral RNA and covers it after virus maturation. NC proteins of lentiviruses contain two zinc fingers that interact with the packaging signal on the viral RNA (Coffin 1992). The role of NC in reverse transcription is to ensure that the tRNA primers bind to the right place on the viral genome and to facilitate the two jumps in the process (Vogt 1997).

### **Pol**

Pol is expressed with Gag when the 9.2 kb MVV viral mRNA is translated (Vigne et al., 1987). The large precursor protein is called Gag-Pol. The subunits of Pol are the reverse transcriptase RT, dUTPase (in non-primate lentiviruses only), integrase IN and protease PR. Upon maturation Pol is cleaved into its subunits by autoproteolytic activity and they are in the capsid core along with the viral RNA. *Pol* is not read in the same reading frame as *gag*. In 5% of transcripts there is a frame shift that ensures that the virions do not contain too much of the enzymes (Luciw and Leung, 1992).

RT is an RNA dependent DNA polymerase that can use RNA or DNA as template. There are 10-20 copies attached to the NC complex in each virion . The C-terminal domain of RT has RNase H activity enabling it to remove RNA from RNA-DNA hybrids. The enzyme is a heterodimer of two RT peptides where the RNase has been removed from one of the molecules. The HIV-1 RT has been described as a hand with the active site in a groove that is protected by “fingers” (Luciw and Leung, 1992; Telesnitsky and Goff, 1997).

IN is at the C-terminal end of the Pol polypeptide and is responsible for the integration of the provirus into the host's genome.

The viral protease PR cleaves the Gag and Pol polyproteins upon virus maturation. PR forms dimers that recognize and cleave a motif of 7-8 hydrophobic amino acids in the polyproteins (Vogt, 1997).

DU or dUTPase is only found in the non-primate lentiviruses. It is located between RT and IN. Its role is to limit the mutation rate by ridding virions of dUTP thereby inhibiting its incorporation into the proviral DNA during reverse transcription. dUTPase appears to be dispensable for MVV infectivity *in vivo* (Petursson et al., 1998).

**Env**

The Env glycoprotein is a precursor of two subunits, TM (trans-membrane) and SU (surface), which make up the spikes that stick out of the lipid bilayer covering the virions. TM of MVV is 44 kDa connected to the Matrix. The 135 kDa SU subunit is connected to TM by non-covalent bonds (Pepin et al., 1998). Each Env “spike” is made up of a trimer of TM proteins connected to a trimer of SU (Vogt, 1997). The env proteins are responsible for receptor binding upon infection.

The SU part of Env contains both variable and conserved domains. The conserved domains are important for receptor binding, but the variable domains change rapidly and allows the viruses to escape recognition by the hosts immune system (Gudnadottir, 1974). Glycosylation also plays a key role in shielding possible antigens from recognition (Haflidadottir et al., 2008; Wei et al., 2003).

**Additional virus proteins****Rev**

The MVV Rev protein is 19 kDa formed in the first phase of transcription by a 1.4 kb mRNA transcript (Sargan and Bennet, 1989). Rev plays a key role in transport of un- or singly spliced mRNA transcripts from the nucleus (Tiley et al., 1990). The N-terminus contains a basic region which binds to a stem-loop structure in the Env RNA called RRE (Rev responsive element) (Tiley and Cullen, 1992; Tiley et al., 1991). The C-terminus contains a leucine rich region which affects trans-activation of transcription (Tiley et al., 1991). Rev contains a nuclear export sequence (Meyer et al., 1996), which facilitates the trafficking of the Rev-mRNA complexes from the nucleus.

**Tat**

MVV Tat was previously described as a trans-activator protein similar to that of other lentiviruses. Recent work has shown that it upregulates the MVV promoter only 2-3 fold as compared to HIV-1 Tat which enhances transcription 100 to 1000 fold under the same conditions. In contrast to the primate lentiviruses, no TAR (trans-acting responsive element) sequence has been identified in the MVV LTR and the MVV Tat binds to cellular transcription factors which target the resulting

complex to an AP-1 binding site in the LTR. The basal promoter activity of MVV is 40 fold higher than that of HIV-1 LTR (Hess et al., 1985). An accessory function of MVV Tat similar to that of Vpr has been suggested, since it is incorporated into virions and induces G2 arrest in the cell cycle of MVV Tat transfected cells (Villet et al., 2003).

### **Vif**

Vif (virion infectivity factor) is found in all lentiviruses except EIAV. The *vif* gene encodes a small (230 amino acids and 29 kDa in MVV, 192 amino acids and 23 kDa in HIV-1) and highly basic protein, separating the *pol* and *env* genes. All lentiviral Vif proteins are tryptophan rich and basic with an isoelectric point (pI) ranging from 9 – 10.8 (Shacklett and Luciw, 1994). The Vif protein of most lentiviruses is cytoplasmic (Goncalves et al., 1994; Goncalves et al., 1995; Michaels et al., 1993; Simon et al., 1997). The Vif protein of FIV has been shown to have both cytoplasmic and nuclear localization, however (Chatterji et al., 2000).

Vif does not have sequence resemblance to known proteins and sequence similarity between the Vif proteins of lentiviruses is rather limited. Despite limited similarity all Vif proteins have a conserved region in the C-terminal half, T/SLQXLA (where X can be any amino acid), which is necessary for their function (Oberste and Gonda, 1992). Vif is expressed in the cytoplasm in the late phase of transcription along with the structural proteins (Volsky et al., 1995).

Vif was implicated in a number of functions, but its precise role in viral infectivity remained elusive for a very long time. The effect of the protein is dependent on the producer cell, i.e.  $\Delta$ Vif viruses produced in permissive cells can cause an infection in non-permissive cells. Vif has been shown to be important for a sustainable *in vivo* infection in MVV, CAEV and SIV (Desrosiers et al., 1998; Harmache et al., 1995; Harmache et al., 1996; Kristbjornsdottir et al., 2004).

### **The $\Delta$ Vif phenotype**

HIV-1 Vif was initially shown to be essential for HIV-1 replication in human lymphocytes and a number of T cell derived cell lines termed non-permissive (CEM, H9 and Hut78), but irrelevant for replication in cells that are called permissive which include some human T cell lines (SupT1, CEM-SS and Jurkat) as well as a number

of non-haematopoietic cell lines such as HeLa, COS7 and 293T (Gabuzda et al., 1992; Strebel et al., 1987). *In vitro* experiments have shown  $\Delta$ Vif MVV grows to some extent in foetal synovial cells (FOS) but poorly in choroid plexus cells (SCP) and macrophages (Kristbjornsdottir et al., 2004). These cells could therefore be termed semi-permissive (FOS) and non-permissive (SCP and macrophages).

*In vivo* experiments (infection of live animals) have been done with FIV in cats (Inoshima et al., 1996), CAEV in goats (Harmache et al., 1996), SIVmac in rhesus macaques (Desrosiers et al., 1998) and MVV in sheep (Kristbjornsdottir et al., 2004). In goats infected with  $\Delta$ Vif CAEV there was very little antibody response (Harmache et al., 1996). In monkeys the  $\Delta$ Vif virus grew very slowly and a very low antibody response was measured (Desrosiers et al., 1998). Sheep infected intratracheally with  $\Delta$ Vif MVV did not get infected, no viruses were isolated from blood or tissues and there was no or very limited antibody response (Kristbjornsdottir et al., 2004).

*In vitro* experiments with  $\Delta$ Vif viruses have been performed with HIV-1 (Gabuzda et al., 1992; Gabuzda et al., 1994; Gibbs et al., 1994; Park et al., 1994; Sova and Volsky, 1993), HIV-2 (Reddy et al., 1995), SIV (Desrosiers et al., 1998), FIV (Lockridge et al., 1999; Tomonaga et al., 1992), CAEV (Harmache et al., 1995) and MVV (Kristbjornsdottir et al., 2004), and Vif turned out to be essential for all of these viruses for a sustainable infection of their natural target cells.

The role of Vif is producer cell dependent.  $\Delta$ Vif viruses from non-permissive cells are non-infectious, whether they are used to infect permissive or non-permissive cells.  $\Delta$ Vif viruses produced in permissive cells are able to infect all cell types. This seems to be the case for all viruses harbouring Vif (Bouyac et al., 1997b; Chowdhury et al., 1996; Courcoul et al., 1995; Gabuzda et al., 1992; Ochsenbauer et al., 1997; Tomonaga et al., 1992; von Schwedler et al., 1993).

The difference of importance of Vif in different cell types can be explained by two possibilities: Either there is an inhibitor in non-permissive cells or there is a factor in permissive cells that is able to perform the role of Vif. This was distinguished by fusing permissive and non-permissive cells. In these heterokaryons the non-permissive phenotype was dominant, implying that the non-permissive cells contained a host factor that was a potent inhibitor of HIV-1 neutralized by Vif and is absent in the permissive cell types (Madani and Kabat, 1998; Simon et al., 1998a).

This elusive host factor was later found to be APOBEC3G, initially termed CEM15 (Sheehy et al., 2002), a member of a recently discovered family of cytosine deaminases (Jarmuz et al., 2002).

The interaction that occurs between Vif and APOBEC3 proteins will be described in a later chapter.

### **Non MVV auxiliary proteins**

In addition to Tat, Rev and Vif HIV-1 encodes for three additional auxiliary proteins called Vpr, Vpu and Nef. As in the case for Vif their principal activity appears to be evasion from an array of cell-mediated antiviral resistance. Vpr facilitates infectivity and causes G2 arrest. Vpu triggers CD4 degradation and enhances virion release by counteracting an antiviral factor named tetherin (Neil et al., 2008). Nef triggers CD4 endocytosis and alters signal transduction in T cells. By limiting the number of CD4 receptors Vpu and Nef reduce the possibility of re-infection of cells and unnecessary interactions with newly formed Env proteins (Malim and Emerman, 2008).

### **Lentiviral infection and replication**

#### **Receptor binding, entry and uncoating**

The first step in lentiviral infection is the connection to the host cell being infected. This depends on the interaction of a receptor molecule on the host cell and the Env glycoproteins of the virus. HIV-1 uses the CD4 molecule and only infects CD4 positive T cells (Dalglish, 1985; Dalglish et al., 1984). HIV-1 also uses chemokine surface molecules as co-receptors, either CXCR4 or CCR5 (Alkhatib et al., 1996). The receptor for MVV is unknown but it is probably a common surface molecule since MVV can enter a very broad number of cell types (Lyall et al., 2000).

When the SU part of Env binds the receptor, a conformational change occurs so that hydrophobic fusion peptides on the N-terminus of TM are uncovered and attach to the cell membrane (Luciw and Leung, 1992). This causes a conformational change in TM and the hairpin trimer is formed. The coat covering the virion is drawn to the cell membrane and fuses to it (Malashkevich et al., 2001).

After the fusion the virus core is launched into the cytoplasm. There the core uncoats and the reverse transcription complex is formed in a process that is not well understood, but cellular factors seem to play a key role (Freed, 2001).

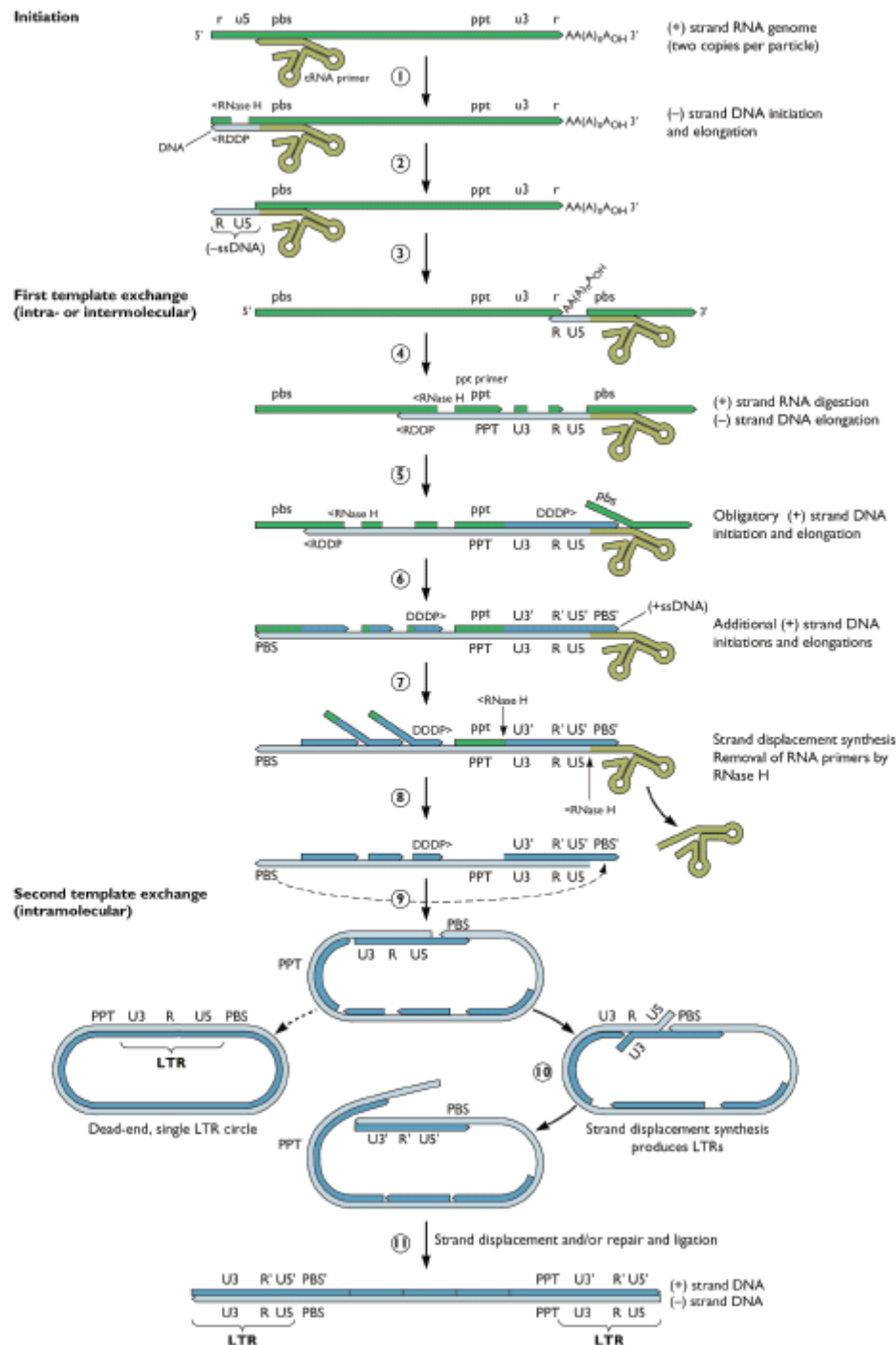
### **Reverse transcription**

The reverse transcriptase RT is a  $Mg^{2+}$  dependent RNA dependent DNA polymerase that lacks proofreading ability (Baltimore, 1970; Lin and Thormar, 1970; Temin and Mizutani, 1970).

Reverse transcription takes place in a number of steps, among them are two jumps where nascent DNA strands are separated from their RNA template and bind to the other end of the virus genome (Figure 4).

1. tRNA primer binds to the primer binding site PBS close to the 5' end of the genome (Sonigo et al. 1985). From the PBS a short –DNA strand is formed to the 5' end. Retroviruses use tRNA primers that they hijack from the producer cell. MVV and HIV use tRNA<sup>lys</sup> (Sonigo et al., 1985). The tRNA is stretched from its original conformation so that 17-18 bases can pair with the PBS (Vogt, 1997).
2. The RNase H activity of RT degrades the RNA from the DNA/RNA hybrid and a DNA strand with a tRNA end is formed. That is called strong-stop –DNA.
3. The 100-150 base strong-stop DNA contains the R part of the LTR and base pairs with the same region at the 3' end of the viral RNA. This first jump is aided by NC and is more often between RNA molecules rather than within the same molecule (Panganiban and Fiore, 1988; Vogt, 1997)
4. The strong-stop –DNA becomes a primer for the formation of the remainder of the minus (-) DNA strand that extends the length of the genome. RNase H degrades the RNA template, but leaves bits at the polypurine sites close to the 5' of the RNA template and close to the 3' end of *pol*. (Freed, 2001; Sonigo et al., 1985)The RNA bits act as primers for the formation of a plus (+) strand DNA which extends until the tRNA primer at the end of the (-) strand has been reverse transcribed and the PBS is formed. Then the tRNA primer is degraded by RNase H.

5. The short (+) DNA strand that is formed is called strong-stop (+) DNA has a PBS that pairs with a similar sequence at the other end of the (-) DNA strand (Freed, 2001). This is the second jump of the process and takes place within a single molecule (Panganiban and Fiore, 1988).
6. The synthesis continues until a double stranded DNA provirus with identical LTR sequences at both ends is formed (Freed, 2001).



**Figure 4.** Reverse transcription. See text for further details. Based on (Katz and Skalka, 1994).

### **Nuclear entry and integration**

Integration of the viral genome into the host cell chromosome is vital both for MVV and HIV (List and Haase, 1997) (Figure 5).

After reverse transcription is completed the proviral DNA is, along with MA, RT, IN and Vpr (HIV-1) in a preintegration complex, PIC (Acheampong et al., 2003). IN cleaves 2 nucleotides off the 3' ends of the genome (Brown, 1997) In MVV sometimes only one nucleotide is cleaved from the end (List and Haase, 1997).

The viral DNA is actively transported into the nucleus. Lentiviruses are transported into the nucleus of non-dividing cells by means of nuclear localization signals NLS in proteins. In HIV-1 MA and Vpr contain the NLS (Acheampong et al., 2003).

After entry into the nucleus the PIC binds to the host chromosome and IN catalyzes a reaction between the 3' OH ends of the provirus and staggered phosphodiester bond in the host chromosome (Brown, 1997). Recent studies show that in lentiviruses this is aided by the host factor LEDGF. Integrase proteins of all lentiviruses including MVV bind directly to LEDGF (Cherepanov, 2007). Finally the repair system of the host cell fills in the nicks at either side of the insertion (Acheampong et al., 2003).

### **Transcription**

Retroviruses use the transcription machinery system of the host cell to express and replicate their genome. The transcription takes place in two phases, in the first phase Tat and Rev (and Nef in HIV) are made but the building blocks and other viral proteins and the genomic RNA are made in the second phase (Mazarin et al., 1988; Sargan and Bennet, 1989).

Transcription is initiated on the junction of U3 and R at the 5' end of the provirus. In HIV-1 TFIID binds to the TATA box and transcription factors like Sp1 and NF-kB bind to promoter sequences (Acheampong et al., 2003). In MVV there are other sites such as AP-1 and AP-4 (Hess et al., 1989) that transcription factors like cJun and cFos bind to (Carruth et al., 1996; Vigne et al., 1994).

Basal transcription from the HIV-1 LTR is low but is transactivated up to a thousand fold by Tat (Freed, 2001; Rabson and Graves, 1997). The MVV LTR basal level of transcription is around 50 fold higher than that of HIV-1.

The RNA transcripts are processed in the same way as other cellular transcripts, with a 5' methyl cap and a polyA tail in U3 on the 3' end (Sonigo et al., 1985).

The Rev protein is necessary to ensure that a subset of the newly formed RNA transcripts are transported un- or single spliced out into the cytoplasm. In the first phase of transcription a fully spliced transcript is formed encoding Tat and Rev which are transported into the nucleus where the second phase is initiated when the level of Rev becomes high enough. Rev binds RRE which is found on all unspliced or partially spliced RNA transcripts. This stops further splicing and transports the transcripts out of the nucleus. Rev traffics between the cytoplasm and the nucleus due to a nuclear localization signal (Acheampong et al., 2003; Freed, 2001; Rabson and Graves, 1997).

### **Expression of viral proteins and virion formation**

Translation of viral proteins takes place on polyribosomes in the cytoplasm, and like with cellular mRNA transcripts, they are either released into the cytoplasm or into the endoplasmic reticulum.

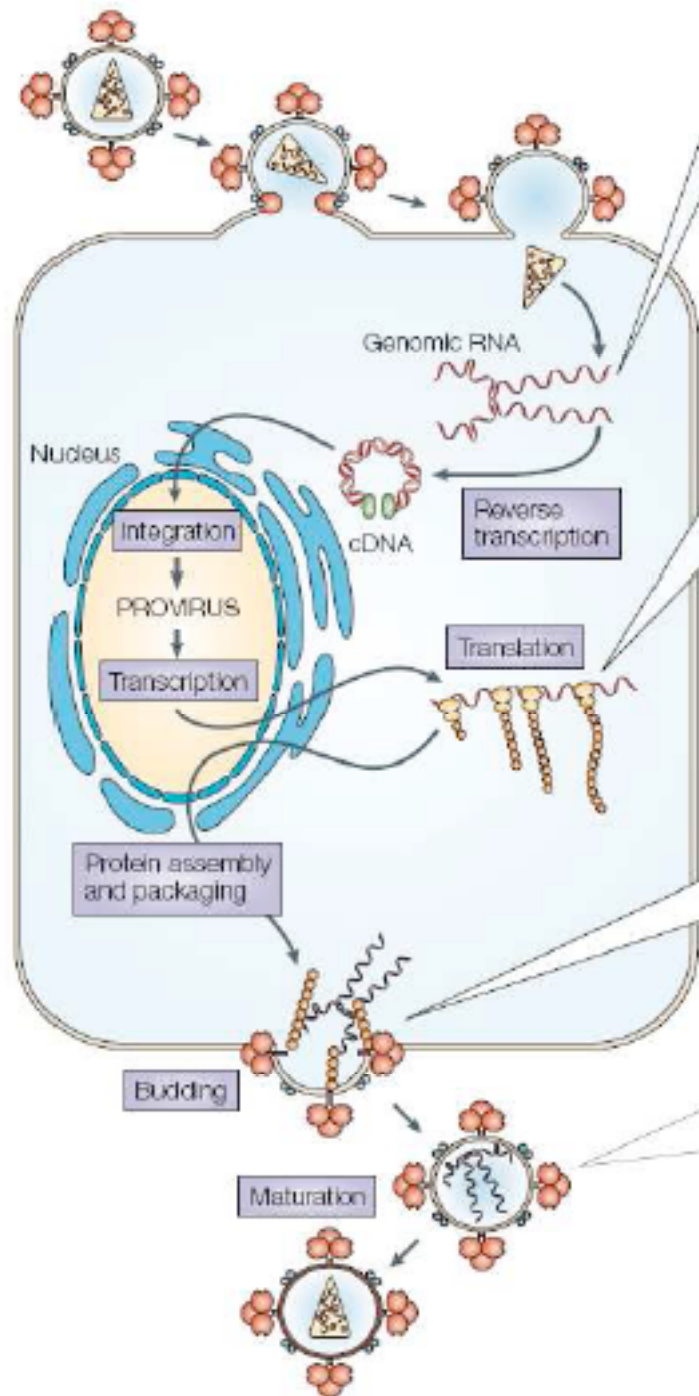
#### **Gag polyproteins**

The N-terminus of MA is cotranslationally modified by myristic acid. This fatty acid modification is essential for Gag membrane binding (Freed 2001, Leciw and Leung 1992). 5-10% of the unspliced mRNA is transcribed into a Gag-Pol polyprotein. Lentiviruses can escape the stop codon at the 3' end of *gag* by ribosomal frame shifting where the ribosomes jump back one nucleotide at the 3' end of *gag* and carry on the *pol* reading frame (Acheampong et al., 2003; Swanstrom and Wills, 1997).

#### **Env**

Env is translated in the same way as membrane bound cellular proteins and is trafficked to the cell membrane through the endoplasmic reticulum (Freed 2001, Swanstrom and Wills 1997). The proteins are cotranslationally moved into the lumen of the endoplasmic reticulum through the use of an N-terminal signal peptide. In the endoplasmic reticulum the Env polyprotein forms intramolecular disulfide bonds, undergoes oligomerization and glycosylation is initiated (Freed, 2001; Vogt, 1997). In the Golgi the proteins are further glycosylated which generates a variety of

products. During Env trafficking through the Golgi Env is cleaved to generate SU and TM which are still linked by non-covalent bonds (Freed, 2001; Swanstrom and Wills, 1997).



**Figure 5.** Lentiviral replication. Schematic of the HIV-1 replication cycle. See text for further details Based on (Paillart et al., 2004).

### **Viron assembly**

Gag, Gag-Pol and Env assemble at the cell membrane. Gag plays a key role in this process; Gag proteins polymerize and form a coat on the inside of the membrane that starts to bud outwards. RNA is incorporated along with other viral proteins along with a number of host factors. In a process called budding an immature virion buds from the cell membrane of the host cell. The viral protease cleaves the Gag polyproteins into their final conformation and an infectious virus is formed.

Gag and Gag-Pol are targeted to the cell membrane through myristylation of the N-terminus of MA . Basic amino acids in MA are targeted to negatively charged phospholipids on the inside of the cell membrane and increase the stability of the interaction of Gag to the membrane (Acheampong et al., 2003; Freed, 2001)

Env is targeted to the cell membrane juxtaposed to assembling virion proteins through an interaction of the cytoplasmic tail of TM and MA . This interaction also protects Env from being endocytosed (Luciw and Leung, 1992; Swanstrom and Wills, 1997).

The packaging of the viral genome is important for the formation of the viral core. Gag alone can form viral like particles (VLPs) *in vitro*, but only if RNA is involved. The assembly process begins with the interaction between RNA and the NC part of Gag (Freed, 2001; Luciw and Leung, 1992; Swanstrom and Wills, 1997). The packaging of the viral genome is specific mediated through a region containing a stem-loop structure called the packaging signal. The region is spliced out of shortened transcripts and therefore they are not packaged (Freed, 2001). Zinc fingers in the NC part of Gag bind to the packaging signal and basic regions surrounding the zinc fingers interact with the RNA (Luciw and Leung, 1992; Swanstrom and Wills, 1997). Each virion contains two copies of the viral RNA genome. The dimer initiation signal is located close to the 5' end of the genome in HIV (Freed 2001), and a dimer initiation signal has also been located in MVV as well as encapsidation determinants which are found in stem loop structures located downstream of the major splice donor (Bjarnadottir et al., 2006; Monie et al., 2005).

The tRNA primer is necessary for the virus and it is packaged through a connection with RT (Luciw and Leung 1992, Swanstrom and Wills 1997).

Budding occurs due to the polymerization of Gag. The C-terminus of CA, the N-terminus of NC and other sequences are important for the process. CA alone can form VLPs, but if NC is expressed along with it, particle formation becomes more

efficient (Freed 2001). Due to the conformation of Gag the protein core formed is roughly spherical. The particles do not pinch off the viral membrane automatically, but rather the pinching off is regulated by regions of Gag, termed L-domains (Freed 2001, Swanstrom and Wills 1997).

### **Maturation**

Virus maturation is visible in electron microscopy as a conformational change of the virus core. Initially the core is spherical, but upon maturation the capsid becomes denser and takes on its final cone shaped formation. This is due to the viral protease cleaving the Gag polyproteins and their structural rearrangement into an electron-dense core (Freed 2001).

The viral protease PR forms homodimers with the active site at the bottom of a cleft between the two subunits. The protease-mediated processing is specific although PR recognises more than one cleavage site (Freed 2001, Luciw and Leung 1992, Swanstrom and Wills 1997). Protease mediated Gag and Gag-Pol processing takes place in an ordered, stepwise cascade of cleavage reactions. This ensures that maturation occurs only after the virus pinches off the producer cells and is regulated by differential affinity of the protease for certain sites (Swanstrom and Wills, 1997).

### **Endogenous retroelements [and retroviruses]**

Darwin could not have foretold that we are descended from viruses as well as from apes. While there is clear evidence that viral diseases affected ancient civilizations, viruses were not defined until the early years of the twentieth century, shortly after the rediscovery of Mendelian genetics. That retroviral genomes can oscillate between cell-free and cell-dependent modes of transmission seemed preposterous before the discovery of reverse transcription in 1970. Those who had earlier provided Mendelian evidence for germ-line transmission of retroviruses were subject to friendly ridicule (Weiss, 2006).

Almost half of the mammalian genome is derived from ancient transposable elements. Two general types, DNA-transposons and retroelements encompass ~2.8% and 42.2% of the human genome, respectively (Deininger and Batzer, 2002; van de Lagemaat et al., 2003). The DNA-transposons jump by a “cut and paste” mechanism without an RNA intermediate, retroelements rely on an RNA transcript that is

retrotranscribed by a reverse transcriptase, or a “copy and paste” mechanism, before being integrated into the genome.

Human retroelements are split into two large groups, the non-LTR and LTR elements. Two of the non-LTR members are present in extremely high copy numbers in the mammalian germ line: the short interspersed elements (SINE) with the prominent Alu and MIR repeats and the long interspersed elements (LINE) containing the autonomous L1 and L2 sequences (Medstrand et al., 2002). SINEs have no protein coding capacity and depend on LINE elements for their amplification. The LTR class elements make up 8% of human chromosomes and include retrotransposons, endogenous retroviruses (ERVs) and repeat elements with HERV origin, such as SINE-R. The SINE-R retrotransposon family contains a partial sequence from a HERV-K LTR. The *env* genes of ERVs confer the potential to spread between cells and individuals. Solitary LTRs of ERVs and retrotransposons, generated by LTR-base homologous recombination processes, are usually one or two orders of magnitude more abundant than preserved or partially complete proviruses (Lower et al., 1996).

None of the HERVs seem to be active and they all appear to have integrated and have been fixed for millions of years. No current transposition activity of HERVs or endogenization of human exogenous retroviruses has been documented so far. However, examples of probable ongoing endogenization processes are known from other mammals. For example, mouse mammary tumour virus (MMTV) and murine leukaemia viruses in mice, Jaagsiekte sheep retrovirus (JSRV) in sheep, porcine endogenous retroviruses (PERV) in pigs, avian leukaemia viruses (ALV) in chickens, and feline leukaemia virus (FeLV) in cats presently have both endogenous and exogenous forms (Boeke and Stoye, 1997). Recently an unprecedented opportunity to capture the endogenization of a retrovirus in real-time has been revealed with the koala retrovirus (KoRV) which currently seems to be in transition between an endogenous and exogenous state (Tarlinton et al., 2008). It has been shown that koala retrovirus (KoRV) is a functional virus that is associated with neoplasia (Tarlinton et al., 2005). KoRV also shows features of a recently inserted endogenous retrovirus that is vertically transmitted. More interestingly, some isolated koala populations have not yet incorporated KoRV into their genomes. Combined with its high level of activity and variability in individual koalas, this suggests that

KoRV is a virus in transition between an exogenous and endogenous element (Tarlinton et al., 2008).

The recent insertion of retroelements in the human genome had led to approximately 1 in every 1000 genetic mutations being caused by Alu insertions (Deininger and Batzer, 1999) and a similar proportion by L1 insertions (Kazazian and Moran, 1998).

A majority of L1 insertions causing disease are on the X chromosome (Kazazian, 2000). The X chromosome has twice the normal level of L1 elements (30% instead of 17%) and therefore it is either a hot spot for L1 insertion or there is a positive selection for L1. It has been proposed that the L1 elements may be highly enriched on the human X chromosome because they may play a role as relays for the X-inactivation signal (Bailey et al., 2000).

Although LTR-retrotransposons comprise around 8% of human DNA, most contain only a single LTR, because of LTR-LTR recombination. Essentially all human LTR-transposons are immobile, although a few HERV elements may remain active. By contrast, the mouse genome harbours active LTR-retrotransposons in the forms of intracisternal A particles (IAPs), MaLR and Etn elements (Baust et al., 2002; Medstrand and Mager, 1998; Ostertag et al., 2003).

L1 elements comprise 17-20% of human and mouse DNA, and are the only autonomous non-LTR retrotransposons in these genomes. Out of the more than 500,000 L1 copies, only around 80-100 are active in the average diploid human genome (Brouha et al., 2003). By comparison, the diploid mouse genome likely harbours ~3000 potentially active L1 elements (Goodier et al., 2001). Although the retrotransposition frequency must be greater in the mouse than in humans, the best estimates of retrotransposition frequency in humans are still relatively high – at least 1 in every 50 sperm (Deininger and Batzer, 2002; Ostertag et al., 2003). The human and mouse genomes also harbour numerous non-autonomous non-LTR retrotransposons, named SINEs. Alu elements are the major SINE in the human genome, whereas B1 and B2 are the major SINE families in the mouse genome. Both B1 and B2 elements are derived from tRNA genes. These SINEs probably all use the proteins encoded by active L1s to mediate their mobility.

The taxonomy of HERVs is a source of confusion. The preferred systematic nomenclature uses the amino acid specificity of the tRNA that hybridizes to the

primer-binding site. The name is defined by adding its one letter code as a suffix to the acronym HERV. HERV-K, for example, uses a lysine specific tRNA as a primer for the initiation of the reverse transcription reaction. The limitations of this approach include the fact that very distantly related viruses use the same tRNAs, and that incomplete information about this short region due to deletions or mutations make classifications of these retroviral sequences almost impossible.

The first HERV-K sequences in the human genome were identified by low stringency Southern blot screenings of human genomic libraries with probes derived from conserved *pol* region from Syrian hamster intracisternal particle A (Ono et al., 1987).

Of the numerous retroviral insertions that have been characterized in vertebrate genomes none have revealed significant homology to exogenous lentiviruses (Gifford et al., 2005; Vogt, 1997). This has been taken as reflecting a recent origin for the lentiviral genus (such that there has not been sufficient time for genome invasion to occur), or a biological barrier to genome invasion, arising either from lentiviral mechanisms for trans regulation of gene expression (Lower et al., 1996) or the lack of specific receptors for these viruses on germ cells (Stoye, 2006). Recently Gifford and colleagues presented evidence for the first endogenous lentivirus, present in the genome of the European rabbit, which was termed rabbit endogenous lentivirus type K or RELIK (Katzourakis et al., 2007). These endogenous lentiviral insertions contain putative *tat* and *rev* genes and corresponding RNA secondary structural motifs demonstrate that germ-line infection by lentiviruses can occur and that lentiviral mechanisms for complex regulation of expression do not preclude successful germ-line colonization. RELIK lacks the *vif* gene. Thus, it is tempting to speculate that the lack of *vif* was in part responsible for the successful host control of the exogenous form of rabbit lentivirus. The sequences of RELIK are so divergent that they preclude a clear-cut examination of the telltale monotonous G-to-A mutational changes that mark the actions of APOBEC3 proteins. The discovery of RELIK extends the host range of lentiviruses to a fifth mammalian order. All the copies of RELIK are defective and analyses indicate that the rabbit genome contains around 25 full-length viruses and 150 solo LTRs. By assuming that copies of RELIK have evolved at the same rate as the rest of the rabbit genome since their divergence, the authors estimated the time of origin of RELIK at least 7 million years, thus

providing the first evidence for an ancient origin of lentiviruses (Katzourakis et al., 2007).

Recently, Lee and Bieniasz reconstructed a proviral clone, termed HERV-K<sub>CON</sub> that resembles the presumed progenitor of HERV-K (HML-2), betaretrovirus-like variants that entered the human genome within the last few million years. Most of the mutations that were reversed to generate the infectious clone were from A to G, suggesting that HERV-K may have been extinguished in humans in part by host defenses that induce mutation of retroviral DNA. Indeed the replication of the reconstituted HERV was inhibited by APOBEC3F, but interestingly neither by APOBEC3G nor human TRIM5 $\alpha$  (Lee and Bieniasz, 2007).

### **Porcine endogenous retroviruses (PERVs) and xenotransplantation**

Xenotransplantation is the transplantation of cells, tissues or organs between different species. In humans, xenotransplantation can include: i) transplantation of solid animal organs, e.g. heart, kidney, liver, ii) transplantation of live animal cells, e.g. pig neuronal and pig pancreatic islet cells, or iii) use of viable animal cells/organs as part of a medical device such as extra-corporeal liver perfusions. The animal species is referred to as the “source” species (Magre et al., 2003).

Successes in allotransplantation, with the availability of haemodialysis and allogenic human organs from donors with brain death, has now led to a shortage of allogenic/human donor organs and tissues, spurring a renewed interest in xenotransplantation.

Pigs have been chosen as a favourable source species for xenotransplantation for a number of logical reasons. Pigs attain sexual maturity within 9 months, have a short gestation period of a 114 days and have large litters of somewhere between 9-16 piglets. Large-scale breeding is highly feasible, no ethical issues or endangered species issues need to be addressed using pigs. Adult pig organ size and life expectancy are compatible with adult humans (Magre et al., 2003).

When transplanted into untreated humans or nonhuman primates, pig organs are rejected hyperacutely within minutes by antibody-mediated complement activation. Hyperacute rejection is the result of incompatibility between donor and recipient encountered in vascularized organ xenotransplantation (Schuurman et al., 2003). Hyperacute rejection is induced by naturally occurring antibodies reactive

against donor antigens (Cramer, 2000; Schuurman et al., 2003; Soin et al., 2000). The major target antigen of human natural xenogenic antibodies is the galactose  $\alpha$ 1,3 galactose sugar residues present on the cell surface of lower mammals and new world monkeys (Schuurman et al 2003).

The birth of the first homozygous  $\alpha$  1,3 galactosyltransferase gene-knockout pigs, not expressing the major xenoantigen recognized by human natural anti-Gal antibodies, was reported in 2003 (Cooper, 2003). The first *in vivo* results following the transplantation of these pig organs into baboons were published in 2004 (Tseng et al., 2004). These newly modified pig organs offered significant progress in terms of graft survival .

The risk of zoonosis, the inadvertent transfer of pathogens between species, is emphasised in xenotransplantation because normal host defences, including the skin and mucosal surfaces are bypassed when human and animal tissue are placed in close contact. Animal pathogens, that under normal circumstances would be non-infectious to humans, may become infectious in a xenotransplantation scenario. The use of immunosuppressive therapy to minimise graft rejection further increases the risk of infection by otherwise non-infectious or latent animal pathogens. The occurrence of zoonotic infections originating from pigs is not unknown. Fatal cases of swine influenza have been attributed to the swine influenza virus and pigs probably played a role in the 1918-19 influenza pandemic (Wells et al., 1991). Encephalitis epidemics in Malaysia and Singapore were caused by the Nipah virus in pigs that recently came from a reservoir in fruit bats (Chua, 2003; Chua et al., 1999; Paton et al., 1999).

Although most known pathogens can be eliminated by SPF “specific pathogen free” breeding and careful screening and monitoring of source animals (Swindle, 1998), the main risk of zoonosis arises from the transfer of source pathogens that are as yet unidentified, difficult to eliminate or maintained in a latent or intracellular state in an asymptomatic host (e.g., endogenous retroviruses such as PERV). Furthermore, genetic modification of source animals to host tolerance induction may alter the host’s susceptibility to animal pathogens or “humanise” animal pathogens allowing them to survive in human recipients (Weiss, 1998).

Viruses are currently under scrutiny as zoonotic agents as they can be efficiently transmitted with viable cellular grafts, may be non-pathogenic in their animal host but could cause serious disease in humans and present a reservoir of potentially untreatable infections. Along with the aforementioned viruses, the swine

influenza virus and the Nipha virus, a number of pig viruses have been linked with the possibility of infecting humans. Among the most notable are members of the *herpesviridae*, Porcine cytomegalovirus (PCMV) (Clark et al., 2003; Fryer et al., 2001) and the Porcine lymphotropic herpesviruses 1 and 2 (PLHV-1 and 2) (Ehlers et al., 1999; Ulrich et al., 1999), Porcine hepatitis E virus (Halbur et al., 2001; Meng et al., 1997; Williams et al., 2001). Porcine encephalomyocarditis virus (EMCV) a picornavirus (Brewer et al., 2001) and the Porcine rotavirus (Santos et al., 1999).

PERV as C-type particles in porcine cell lines were first described in the 1970s (Breese, 1970). These particles were produced from a number of pig kidney cell lines, among them PK-15 (Armstrong et al., 1971). Viral RNA was found to hybridise to DNA but not RNA extracted from pig liver, indicating that despite the apparently endogenous nature of these viruses they were not actively transcribed in the pig liver (Todaro et al., 1974).

With the revival of xenotransplantation in the 1990s there has been renewed interest in the biology of PERVs. In 1997 Clive Patience and colleagues demonstrated the ability of PERVs to infect human cells *in vitro* (Patience et al., 1997). In this study virus spontaneously released in PK-15 (PERV-PK) cell supernatant was found to infect porcine ST-IOWA, mink lung (Mv-1-Lu) and human embryonic kidney (293) cell lines, while MPK virus (PERV-MPK) was restricted to ST-IOWA cells. Upon co-cultivation, PERV-PK was able to infect a broader range of human cell lines. While RT activity was evident in PERV-PK infected 293 cell supernatants, indicating productive infection, its absence in a PERV-PK infected human fibrosarcoma (HT1080) cell line and primary human PBMCs, despite the presence of PERV-PK sequences, suggested non-productive infection. Southern blot analysis indicated multiple copy numbers of PERV in DNA extracted from the heart, kidney and spleen of domestic pigs, confirming their endogenous nature (Patience et al., 1997).

To date, three subgroups of PERV have been identified, termed PERV-A, -B and -C, which contain highly conserved *gag* and *pol* genes, but diverge in their envelope sequences. The observation that full-length PERV sequences (approximately 8-15) vary among swine strains (Akiyoshi et al., 1998) led to the idea that selective outbreeding of source pigs that expressed less PERV may reduce the proviral load in source organs, thereby reducing the possible risk of transmission to humans. Screening of SPF pigs for their genomic distribution of PERV showed the

absence of PERV-C in some animals. However, persistently high levels of full-length PERV genomes were found in organs of seven PERV-C negative animals. Isolation and envelope sequencing of these full-length transcripts identified both functional and defective PERV-A and PERV-B like envelope sequences which may indicate replication competence of these transcripts (Bosch et al., 2000). In another study, the presence of PERV mRNAs in a panel of tissues from SPF and conventional pigs was determined by a semi-quantitative RT PCR. The results showed that PERV mRNA levels were the same in both SPF and conventional pig tissues. PERV-A and PERV-B envelopes were more represented compared with PERV-C envelopes and PERV expression varied in tissues with the pancreas expressing the least and kidneys expressing the highest levels of PERV mRNA (Clemenceau et al., 1999). Furthermore, isolation of three PERV-A types and one PERV-B type replication – competent molecular clones, from a PK-15 cell line genomic library may suggest that the pig genome harbours a limited number of infectious PERV-A and PERV-B sequences (Krach et al., 2001). Identification of such replication-competent retroviruses in the pig genome may also help in identifying pig breeds which produce lower levels of PERV or are devoid of individual proviruses as a result of polymorphisms, thereby increasing the possibility of cloning PERV-free pigs (Niebert et al., 2002; Oldmixon et al., 2002). However, the mere presence of PERC, even if non-infectious, could lead to recombination events, mutation and other possible changes, which may alter their “non-pathogenic” state. Non-infectious, replication-deficient PERV may recombine with replication-competent, infectious PERV, or other related endogenous retroviruses. One example of a recombinant PERV is already known. The PERV-NIH isolate contains an *env* sequence with the receptor-binding domain derived from PERV-A and the cell fusion domains derived from PERV-C (Wilson et al., 2000).

PERV has been shown to be able to infect a panel of human cells (Patience et al 1997). Productive infection of primary human endothelial and mesengial cells was shown by detection of RT activity from the supernatants of cells co-cultivated with virus producing cells (Martin et al., 2000). The danger of PERV in xenotransplantation was further emphasised when mitogenic stimulation of primary porcine PBMCs from two pig strains, the National Institutes of Health (NIH) miniature pig (minipig) and the Yucatan pig, was found to release porcine C-type retroviruses capable of productively infecting pig and human cell lines. The *pol* gene

of these viruses showed 95% nucleotide identity with PERV-PK virus (Wilson et al 1998). Xenotransplantation-relevant porcine cells, including primary porcine aortic endothelial cells (PAECs) form the the main interface between a xenograft and a recipients leukocytes and tissues), hepatocytes, lung and skin from various pig strains and breeds were also shown to express PERV mRNA. These PAECs released infectious PERV, which transferred to human 293 cells upon co-cultivation with expression of both PERV-A and PERV-B subgroups (Martin et al 1998).

The ability of certain PERV to infect human cells has been documented *in vitro* (Martin et al., 2002). However, using the same *in vitro* conditions, it has not been possible to achieve viral transmission from cells from selected strains of miniature swine (Martin et al 2002), and PERV transmission into humans or nonhuman primates has never been observed *in vivo* after exposure to living porcine tissues (Irgang et al., 2003; Magre et al., 2003). These discrepancies between *in vitro* and *in vivo* results could be explained by the presence of natural immunity against PERV in human and/or primates (Blusch et al., 2002).

### **APOBEC3 protein family of cytosine deaminases**

#### **The APOBEC3 family cytosine deaminases**

APOBEC3 (apolipoprotein B mRNA-editing catalytic polypeptide-like 3) proteins belong to a family of polynucleotide cytosine deaminases that includes AID (activation induced deaminase), APOBEC1 (A1), APOBEC2 (A2) and APOBEC4 (A4) (Conticello et al., 2005; Jarmuz et al., 2002; Rogozin et al., 2005). These proteins catalyze the deamination of cytosine to uracil in single-stranded DNA and/or RNA affecting diverse physiological functions. All family members have either one or two deaminase motifs (Z domains) comprising the sequence His-Xaa-Glu-Xaa<sub>23-28</sub>-Pro-Cys-Xaa<sub>2-4</sub>-Cys (where X can be nearly every amino acid), the histidine and two cysteine residues coordinate a Zn<sup>2+</sup> ion, and the glutamic acid residue is involved in proton shuttling. Catalysis results in hydrolytic deamination at the C4 position of the cytosine base, thereby converting cytosine to uracil in a process called editing (Teng et al., 1993; Wedekind et al., 2003).

APOBEC1 was the first family member to be identified and therefore the whole family gets its cumbersome name from its function. APOBEC1 is involved in lipid

metabolism (Teng et al., 1993). It is expressed in gastrointestinal tissues and specifically edits C6666 in apolipoprotein B mRNA to generate a premature stop codon that leads to a truncated functional form of apolipoprotein B (Teng et al., 1993; Yamanaka et al., 1995). The longer apoB100 and the shorter apoB48 have different effects on cholesterol metabolism (Anant and Davidson, 2001).

The second family member AID is expressed in B cells and is essential for three types of immunoglobulin-gene diversification: somatic hypermutation, class switch recombination and gene conversion through C to U editing at the immunoglobulin loci (Eto et al., 2003; Muramatsu et al., 2000; Muramatsu et al., 1999; Petersen-Mahrt et al., 2002). These processes are an integral part of the DNA-level modifications that drive maturation of the vertebrate antibody response to pathogens. Recently AID proteins from a panel of different vertebrates were shown to be able to inhibit the replication of the retroelements LINE1 and MusD. AID was able to inhibit the replication in a deamination independent manner. AID was as well shown to be expressed in reproductive tissues in mice, which might suggest that AID has or may have had innate immune functions in addition to its role in creating antibody diversity. This data furthermore pinpoints AID as the ancestor of the modern day APOBEC3 proteins (MacDuff et al., 2009).

APOBEC2 is expressed most prominently in cardiac and skeletal muscle and is regulated by NF- $\kappa$ B in human hepatocytes (Liao et al., 1999; Matsumoto et al., 2006; Prochnow et al., 2007), but it has failed to show activity on cytosine with DNA or RNA (Harris et al., 2002; Liao et al., 1999). APOBEC4 is the most recently defined member of the family; it is primarily expressed in testes and is most homologous to APOBEC1 of the other family members (Rogozin et al., 2005). The physiological functions of APOBEC2 and APOBEC4 are still unknown.

APOBEC3 proteins have come to the forefront of this family by their association with HIV-1 and their role not only in antiretroviral defenses but also their activity against mobile genetic elements e.g., (Chiu and Greene, 2008; Cullen, 2006; Harris and Liddament, 2004; Stenglein and Harris, 2006). The APOBEC3 genes are syntenically arrayed between two conserved flanking genes, CBX6 and CBX7 (LaRue et al., 2008). Uniquely among the family, APOBEC3 genes vary in number among species, from one in rodents to seven in primates. In humans the APOBEC3 genes are clustered on chromosome 22 (Harris and Liddament, 2004; Jarmuz et al., 2002) and have been given names in alphabetical order (APOBEC3A to

APOBEC3H). Early analyses suggested that APOBEC3E was a pseudogene, but now it is thought that APOBEC3D and APOBEC3E form the N- and C-terminal regions of one double-domain protein, called APOBEC3DE (Dang et al., 2006). Four of these genes have two deaminase domains (APOBEC3B, APOBEC3DE, APOBEC3F and APOBEC3G), and the others (APOBEC3A, APOBEC3C and APOBEC3H) contain a single deaminase domain like the other members of the APOBEC family.

The APOBEC3 proteins all appear to be globular proteins centered upon the His-Xaa-Glu-Xaa<sub>23-28</sub>-Pro-Cys-Xaa<sub>2-4</sub>-Cys conserved motif (Harjes et al., 2009). The human proteins that contain two catalytic domains (APOBEC3B, APOBEC3DE, APOBEC3F and APOBEC3G in humans) the entire unit is duplicated to form two globular structures with a catalytic globular C-terminal half and a pseudocatalytic globular N-terminal half (Harjes et al., 2009).

Each Z domain clearly belongs to one of the three distinct phylogenetic clusters, termed, Z1, Z2 and Z3 (LaRue et al., 2009; LaRue et al., 2008). Z-domain assignments can be made simply by scanning predicted polypeptide sequences for key identifying residues. This determination is facilitated by the fact that the Z domain of all known A3 genes is encoded by a single exon. For instance, Z1 domains have a unique isoleucine (I) adjacent to a conserved arginine common to all DNA deaminases (Chen et al., 2008). Z2 domains possess a unique tryptophan-phenylalanine (WF) motif five residues after the (pseudo)catalytic glutamate. Finally, Z3 domains have a TWSPC<sub>x2-4</sub>C zinc-coordinating motif, whereas both the Z1 and Z2 domains have a SWS/TPC<sub>x2-4</sub>C motif. Since many A3 proteins have been subject to positive selection (Sawyer et al., 2004), this Z-based scheme is also substantially more robust to evolutionary constraints and pressures that have acted (and continue to act) on A3 proteins in different lineages.

APOBEC proteins show a distinctive preference for deaminating cytosines in local sequence contexts. For instance, human APOBEC3G prefers the second C in CpC dinucleotides (GG dinucleotides in the viral plus strand cDNA), whereas human APOBEC3F and APOBEC3B have a preference for the C of TpC dinucleotides (GA in the viral cDNA) (Bishop et al., 2004; Doehle et al., 2005a; Harris et al., 2003b; Liddament et al., 2004; Wiegand et al., 2004). It has been shown for human APOBEC3F and APOBEC3G that the C-terminal deaminase domain confers the local target site preference. Mouse APOBEC3 shows a joint preference for TpC

dinucleotides (~60%) and CpC (~30%). Cow and Sheep APOBEC3Z2-Z3 proteins both have a dinucleotide preference similar to that seen for the mouse APOBEC3 protein (Jonsson et al., 2006), paper II. The recently cloned equine APOBEC3Z2a-Z2b prefers the C of TpC dinucleotides (Bogerd et al., 2008). Almost all of the APOBEC3 proteins tested have shown a preference for a pyrimidine (Y) base in the -1 position. In contrast and notably, the porcine APOBEC3Z2-Z3 protein preferred GpC in nearly half of the sequences analyzed (Jonsson et al., 2006), paper II, this preference is shared by both human APOBEC3DE (Dang et al., 2006) and feline APOBEC3Z2-Z3 (formerly A3CH) as the only APOBEC3 proteins shown to prefer a purine base in the -1 position, which is what AID prefers (Beale et al., 2004). In some cases a high sequence preference extends to the -2 position, the artiodactyl APOBEC3Z2-Z3 proteins have a high preference for T as does the mouse APOBEC3 protein (Jonsson et al., 2006), paper II.

It has been difficult to resolve the structure of the APOBEC3 proteins, since these are highly insoluble and prone to precipitation during purification (Chelico et al., 2006; Iwatani et al., 2006). Recently, Chen et al. determined the structure of the APOBEC3G catalytic domain, revealing six  $\alpha$ -helices, including two that form the zinc-coordinating active site, arranged over a hydrophobic platform consisting of five  $\beta$ -strands (Chen et al., 2008).

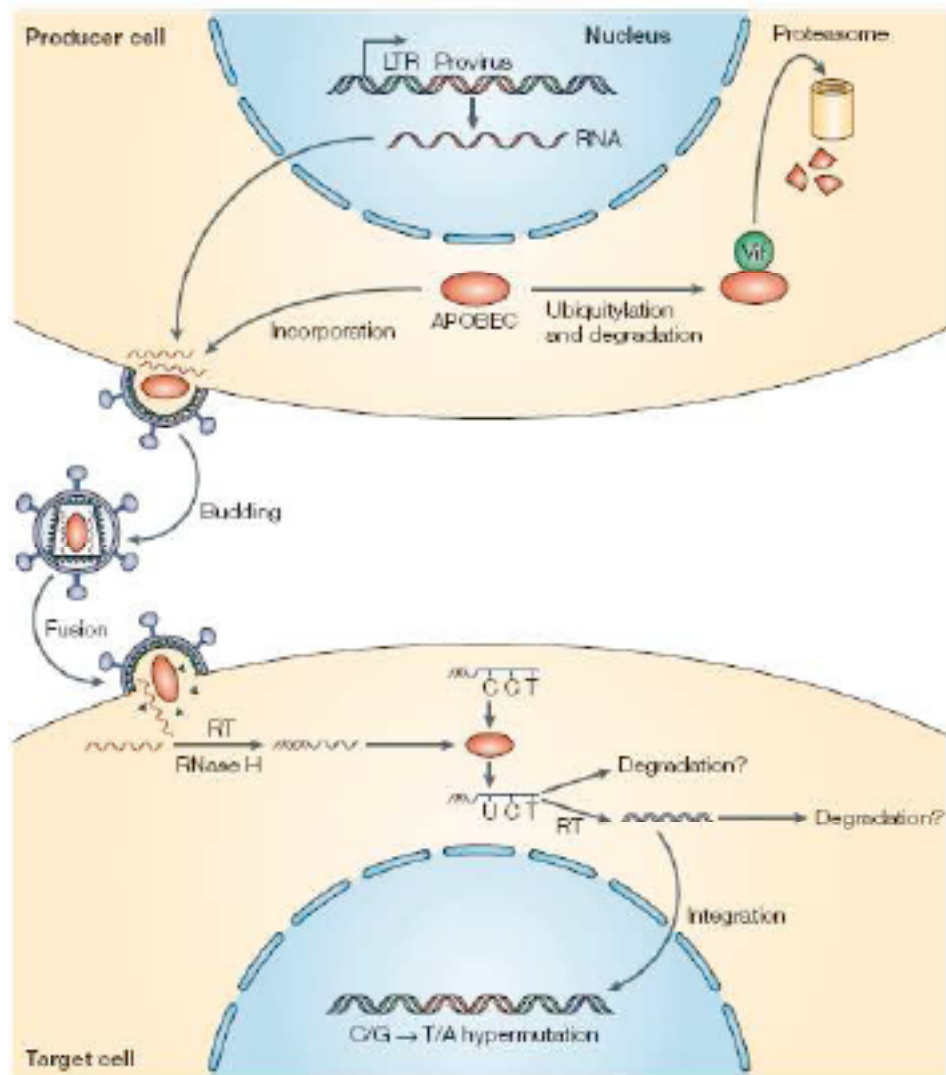
In the cytoplasm of H9 T cells and mitogen activated CD4 T cells APOBEC3G is assembled in 5-15 MDa high-molecular mass (HMM) ribonucleoprotein (RNP) complexes and the deoxycytidine deaminase activity of APOBEC3G is greatly inhibited in these complexes. The HMM A3G complexes can be artificially converted to an enzymatically active low-molecular mass (LMM) form by treatment with RNase A, suggesting that RNA components play an important role in the assembly of HMM A3G complexes (Chiu et al., 2005).

APOBEC3 proteins probably use a well conserved way of engaging retroelements as APOBEC3 proteins have been shown to be active against elements both in mammals, lower vertebrates and even in such distantly related organisms as yeast (Dutko et al., 2005; Schumacher et al., 2005).

### **Inhibition of exogenous and endogenous retroelements**

Due to the position of APOBEC3G in a family of cytosine deaminases the obvious question was whether it had the ability to mutate HIV-1 DNA, since it had been shown to possess DNA mutational activity in an *E. coli* based assay system (Harris et al., 2002). It was then shown that expression of APOBEC3G in virus producing cells yields APOBEC3G containing progeny virions that, upon infecting fresh cells, produced reverse transcripts with numerous guanosine (G) to adenosine (A) transition mutations on the plus (proviral cDNA) strand (Harris et al., 2003b; Mangeat et al., 2003; Mariani et al., 2003; Zhang et al., 2003).

For APOBEC3G to be able to inhibit HIV-1, it needs to be packaged into nascent progeny virions. During encapsidation, APOBEC3G interacts with the Gag polyprotein and with viral RNA. It is thought that APOBEC3G is packaged through specific interactions between the N-terminal deamination domain and the NC region of Gag as well as with binding to RNA (Bogerd and Cullen, 2008; Khan, 2005; Luo et al., 2004; Navarro and Blackwell, 2005; Schafer et al., 2004; Svarovskaia et al., 2004; Zennou et al., 2004). It is as of yet unresolved which type(s) of RNA it is that promotes the packaging of APOBEC3G, though it is believed that HIV-1 genomic RNA and 7SL RNA are likely candidates (Khan et al., 2007; Khan et al., 2005; Luo et al., 2004; Wang et al., 2007). Amino acids 124 and 127 in APOBEC3G play a key role in RNA binding and mutants of those residues yield proteins that do not package and therefore do not inhibit infection (Huthoff and Malim, 2007; Wang et al., 2007; Zhang et al., 2007) (Figure 6). Several lines of evidence point towards that the deaminase activity of human APOBEC3G is required for efficient inhibition of Vif defective HIV-1 as well as the two retrotransposons Ty1 and the MusD (Miyagi et al., 2007; Schumacher et al., 2008)



**Figure 6.** APOBEC3 mediated inhibition of lentiviral infection. APOBEC3 (red) is produced in the producer cell and is incorporated into the budding virion, Vif (green) can reduce or eliminate APOBEC3 by targeting it for proteosomal degradation. Should APOBEC3 escape Vif and reach the target cell, it can deaminate C to U in the –DNA strand, resulting in G to A hypermutations that affect virus viability. Uracil residues have been hypothesised to trigger degradation of the retroviral DNA before it can integrate into the host genome (Based on Harris and Liddament 2004).

### Deamination-independent inhibition

Recent work has shown that APOBEC3G can also inhibit infection somewhat without the aid of G to A hypermutations. It has been shown by mutating any of the catalytically important residues in the C-terminal deamination domain that editing is entirely mediated by the C-terminal domain (Hache et al., 2005; Newman et al.,

2005). It was then shown that these mutants could still inhibit HIV-1 infection to a significant extent (Newman et al., 2005). It is still unknown what the underlying mechanism is but it is becoming widely accepted that wild-type and catalytically inert APOBEC3G are able to diminish the accumulation of reverse transcripts in newly infected cells (Mangeat et al 2003, Mariani et al 2003, Bishop et al 2006, Guo et al 2006, Holmes et al 2007, Luo et al 2007, Mbisa et al 2007, Bishop et al 2008) (Bishop et al., 2006; Bishop et al., 2008; Guo et al., 2006; Holmes et al., 2007; Luo et al., 2007; Mangeat et al., 2003; Mariani et al., 2003; Mbisa et al., 2007).

Apart from HIV-1, a wide array of retroviruses (and members of other virus families) have been tested to see whether APOBEC3 proteins have an effect on their replication, it can be said that most if not all retroviruses tested are sensitive to at least one APOBEC3 protein. An emerging theme is that although the APOBEC3 proteins seem to be a broad defense against retroviruses, each retrovirus seems to have adapted and has come up with a strategy that is specific for evasion of the APOBEC3 proteins of its host (see tables 1 and 2 and the next chapter).

	Retroviruses										Other viruses			
	HIV-1		SIV <sub>mac</sub>		SIV <sub>agm</sub>		HTLV	EIAV	MLV	PFV ΔBet	HBV	AAV	Vac	HPV
	WT	Δvif	WT	Δvif	WT	Δvif								
A3A	-	-	-	-	-	-						+		+
A3B	+	+/-	+	+	+	+	+/-		+/-		+	-		-
A3C	-	+/-	-	+	-	+	+/-		-	+	+/-	-		+
A3DE	-	+/-	-	+	+	+			-					
A3F	-	+	-	+	+	+	+/-		-	+	+	-	-	
A3G														
Human	-	+	-	+	+	+		+	+	+	+	-	-	
Chimp	-	+	-	+	+	+				+				
MAC	+	+	-	+	-	+								
AGM	+	+	-	+	-	+								
A3H	-	-	-	-	-	-			-				-	+
MmA3	+	+	+	+	+	+			-	+				
BtAF	+	+							+					
OaA3F	+	+							+					
SsA3F	+	+							+					
FcA3C	+	+								+				

**Table 1.** Summary of restriction patterns by APOBEC3 proteins. Table showing the inhibitory effect of APOBEC3 proteins of a range of viruses and the species specificity of the Vif:APOBEC3 interaction (+ denotes ability to inhibit virus whereas – denotes inability to inhibit virus).

### **Modes of viral escape**

As mentioned earlier, all lentiviruses except EIAV code for Vif, which is a potent APOBEC3 antagonist. The main focus of research has been on the HIV-1 Vif-APOBEC3G interaction, which was the initial observation giving the clue to the role of Vif (Sheehy et al., 2002). Vif induces polyubiquitylation of A3G for subsequent degradation in the proteasome, thus preventing the incorporation of A3G into the viral particle (Conticello et al., 2003; Mariani et al., 2003; Sheehy et al., 2003; Yu et al., 2003). Vif achieves this by acting as an adaptor connecting A3G and the cullin5-ElonginB-ElonginC-Rbx ubiquitin ligase (Yu et al., 2003). The conserved SLQXLA motif in the C-terminus of Vif interacts with ElonginC and Elongin B and is referred to as the BC box (Yu et al., 2003; Yu et al., 2004). A zinc-binding motif  $Hx_5Cx_{17-18}Cx_{3-5}H$  that spans Vif amino acids 108-139, which is conserved within the primate lentiviruses, together with a downstream  $LPx_4L$  motif mediate Cul5 interaction (Stanley et al., 2008; Zhang et al., 2008).

The interaction between A3G and Vif has been mapped to amino acids 128-130 in A3G and to the N-terminus of Vif (Huthoff and Malim, 2007; Russell and Pathak, 2007; Schrofelbauer et al., 2004). In addition to APOBEC3G, A3F, A3C and A3DE are also counteracted by Vif, and different sections of HIV-1 Vif mediate interaction with the various APOBECs. Thus, residues at positions 22 to 44 are important for A3G binding but not A3F, and conversely, residues 11 to 17 and 74 to 79 are important for A3F but not A3G binding. A region in Vif harboring hydrophobic amino acids (positions 55-72) is required for both A3F and A3G binding. The same residues that are important for A3F binding and degradation are also important for A3C and A3DE (Zhang et al., 2008). It has also been reported that Vif mediates degradation of A3G by the proteasomal pathway in the absence of polyubiquitylation (Dang et al., 2008).

The Vif-mediated APOBEC3 restriction is species specific and may therefore be one of the factors that determine virus host range and cross-species transmission (Jonsson et al., 2006; Simon et al., 1998b) (table 1).

Non-lentiviral retroviruses that do not have a Vif protein have come up with novel mechanisms to evade the APOBEC3 proteins they are challenged with.

Murine leukaemia virus (MLV) is insensitive to inhibition by mouse APOBEC3. The evasion seems to take place by viron exclusion (Abudu et al., 2006; Doehle et al., 2005b; Kobayashi et al., 2004) and a fi mechanism has as well been reported in

the cleavage of the mouse APOBEC3 protein by the viral protease (Abudu et al., 2006). HTLV-1, which like HIV-1 targets CD4<sup>+</sup> T cells, appears mostly resistant to APOBEC3G and seems to exclude APOBEC3G from the virions through an acidic extension of the virus nucleocapsid (Derse et al., 2007).

Foamy viruses encode a protein called Bet that binds to the APOBEC3 protein and prevents it from being packaged into the nascent virions. Unlike Vif, Bet does not target the APOBEC3 proteins for proteasomal degradation but rather binds the APOBEC3 protein and sequesters it away from the site of virus formation (Delebecque et al., 2006; Russell et al., 2005).

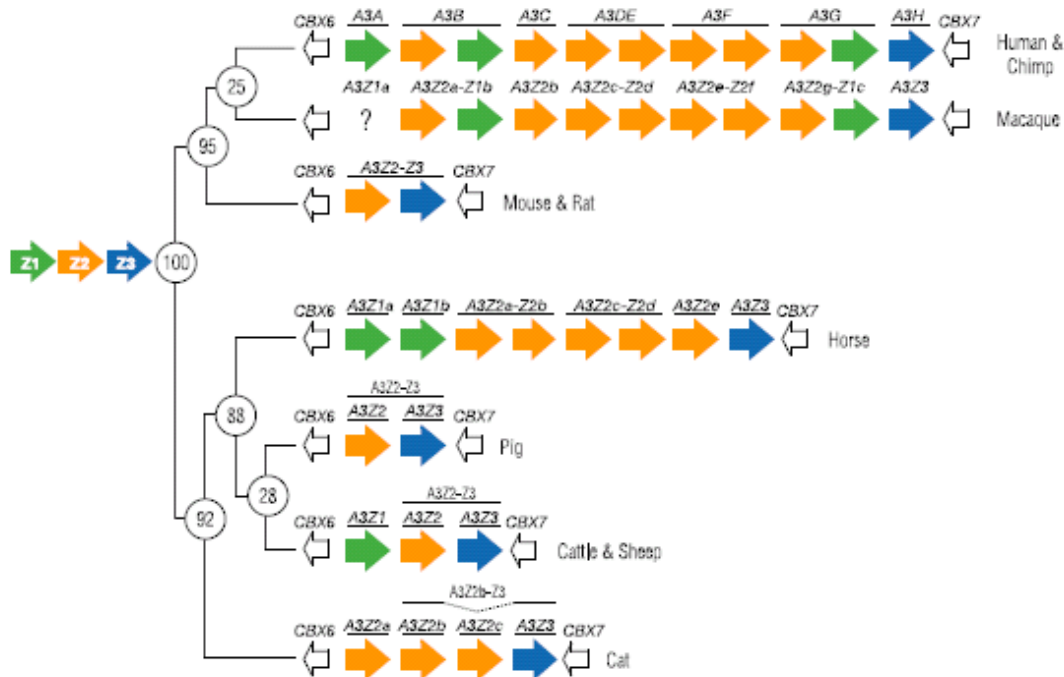
	Expression profile	Subcellular localization	LTR Retroelements			Non-LTR Retroelements	
			IAP	MusD	Ty1	L1	Alu
A3A	Keratinocytes, spleen, small intestine, monocytes, macrophages, colorectal adenocarcinoma, Burkitt's lymphoma, chronic myelogenous leukemia	N/C	+	+		+	+
A3B	Keratinocytes, colon, small intestine, testis, ovary, stem cells, lung carcinoma, colorectal adenocarcinoma, Burkitt's lymphoma, chronic myelogenous leukemia	N	+	+		+	+
A3C	Many tissues and a variety of cancer cell lines	N/C	+	+	+	+	+
A3DE	Many tissues	N/C				-	
A3F	Many tissues and probably coexpressed with APOBEC3G	C	+	+	+	+/-	+
A3G	Primarily expressed in the lymphoid and myeloid cell lineage, thymus, tonsil, bone marrow, also observed in testis, ovary, uterus, brain, heart, lung, liver, kidney, spleen, pancreas, colorectal adenocarcinoma, and Burkitt's lymphoma	C	+	+	+	-	+
Human			+	-		-	
Chimp			+	-		-	
MAC			+	-		-	
AGM			+	+/-		+/-	
A3H	Peripheral blood lymphocytes, testis, ovary, fetal liver, skin, cerebellum, colon, small intestine	N/C				-	
MmA3		C	+	+	+	-	
BtA3F	Peripheral blood lymphocytes, fetal tissues	C					
OaA3F	Peripheral blood lymphocytes, macrophages, fetal tissues, choroid plexus	C					
SsA3F	Peripheral blood lymphocytes, fetal tissues	C					

**Table 2.** Summary of restriction patterns and subcellular localization of APOBEC3 proteins. Table showing the inhibitory effect of APOBEC3 proteins of a range of endogenous retroelements and subcellular localization of the APOBEC3 proteins. N=nuclear C=cytoplasmic (+ denotes ability to inhibit virus whereas – denotes inability to inhibit virus).

### APOBEC3 evolution and positive selection

One of the most interesting things about the APOBEC3 family is the vast difference in gene numbers between mammalian species. It was first noticed by comparing the human and rodent loci, and with more genomes being sequenced, the picture becomes more and more intriguing. As shown in paper III sheep and cows have three APOBEC3 genes and pigs have two (L... et al., 2008). Other recent projects have shown the same sort of variety with the cat locus consisting of four APOBEC3 genes

(Munk et al., 2008) and horses having 6 APOBEC3 genes (Bogerd et al., 2008) (Figure 7).



**Figure 7.** Schematics of the APOBEC3 repertoires of mammals whose genomes have been sequenced. Z1, Z2, Z3 domains are shown in green, orange and blue, respectively. The numbers at the phylogenetic branch points indicate the approximate time, in millions of years, since the divergence of the ancestors of the clades of the indicated present day species (from LaRue et al 2009).

The APOBEC3 gene family seems to be placental mammal specific since the sequencing of the genomes of representative species of monotremata, the duck-billed platypus (Warren et al., 2008) and marsupial, the opossum (Mikkelsen et al., 2007) revealed no APOBEC3 genes. It will be interesting to see what future sequencing projects bring as there is still missing information from some of the placental mammal super-orders such as *Afrotheria* and *Xenarthia*.

The current day human APOBEC3 locus of 7 genes with 11 Z domains seems to have expanded from the probable 3 domain Z1-Z2-Z3 formation of the placental mammal ancestor during the early primate evolution about 25 – 65 million years ago. A minimum of 8 recombination events were likely required to generate the present day primate APOBEC3 locus (LaRue et al., 2008).

The human APOBEC3 locus seems quite unstable and two rather common polymorphisms have been described recently showing that the APOBEC3 proteins are likely to be still engaged in considerable genetic conflict.

Kidd and coworkers presented a 29.5kb common deletion that removes the whole of the APOBEC3B gene. Characterization of 1277 human samples with a PCR-based genotyping assay found that this deletion varied greatly among major ethnic or continental groups. The deletion is rare in Africans and Europeans (frequency of 0.9% and 6%), more common in East Asians and Amerindians (36.9% and 57.7%), and almost fixed in Oceanic populations (92.9%). Overall the deletion has a worldwide frequency of 22.5% (Kidd et al., 2007).

Two independent mutations in APOBEC3H give rise to an unstable protein that has no detectable antiviral activity, the inferred ancestral APOBEC3H protein is active against HIV but sensitive to HIV-1 Vif and very active against LINE-1 (OhAinle et al., 2008). The stability of the protein was lost twice during recent human evolution through the acquisition of two different amino acid mutations that decrease the half-life of the protein. Active APOBEC3H alleles are found among 51-52% of African Americans and members of the Yoruba people from Nigeria, 10-18% in Europeans and European Americans and only 3-4% in East Asians and Chinese-American populations (OhAinle et al., 2008). Despite the fact that Z3 (or APOBEC3H-like) genes have been found in all mammals that have been subjected to research, the Z3 domain is the only one that has not undergone any sort of duplication event and is always found on the distal (or 3') end of the APOBEC3 locus next to the CBX7 gene (OhAinle et al 2008, LaRue et al 2008 I, LaRue et al 2008 II – paper III).

APOBEC3G and the other members of the APOBEC3 family in primates, except APOBEC3A which is under purifying selection, seem to have been engaged in a genetic conflict throughout the history of primate evolution or at least during the last 33 million years (Sawyer et al., 2004). The positive selection seems to be much more ancient for it to have been only caused by the relatively modern lentiviruses. This observation is further enhanced by the fact that this positive selection is evident in primate species regardless of whether they are subject to SIV infection or not (Zhang and Webb, 2004).

## 2. AIMS OF THE STUDY

The overall aim of the thesis was to investigate innate host defenses against retroviruses, and specifically the APOBEC3 proteins of artiodactyl species sheep, cattle and pigs.

The specific aims of the thesis were:

- To characterize a maedi visna virus mutant with attenuated replication in sheep macrophages and in vivo (paper I).
- To clone and characterize APOBEC3 from sheep and other artiodactyls and investigate the evolution of these proteins (papers II and III).
- To investigate whether cross-species expression of an APOBEC3 protein can impede virus transmission by stably expressing human APOBEC3G in a pig cell-line and monitoring whether its expression inhibits the transmission of PERV from pig to human cells (paper IV).

### **3. MATERIALS AND METHODS**

## **Cells and cell-culture**

### **Macrophage isolation and culture**

Macrophage cultures were established as follows. Heparinized blood (100 - 400 ml) was collected from normal sheep, and peripheral blood mononuclear cells obtained by sedimentation on Histopaque-1077 (Sigma) were washed repeatedly in phosphate-buffered saline and resuspended at  $12 \times 10^6$  cells/ml in growth medium supplemented with  $5 \times 10^{-5}$  M mercaptoethanol. They were then seeded into T25 tissue culture flasks (25cm<sup>2</sup>, Nunc Inc.) or into 6-well plates. After incubation at 37°C in a humidified atmosphere of 5% CO<sub>2</sub> for 24h, supernatant and unattached cells were removed, the slide was washed twice with phosphate-buffered saline, and 5 ml new growth medium was added to each of the culture flasks. Adherent cells were further incubated for at least 7 days before they were infected.

### **Virus and cells**

The molecularly cloned viruses KV1772 (formerly KV1772kv72/67; GenBank accession number S55323.1) and KS1 (formerly LV1-1KS1; GenBank accession number M6069.1) have been described previously (Andresson et al., 1993; Skraban et al., 1999; Staskus et al., 1991). Sheep choroid plexus (SCP) cells established as described previously (Sigurdsson et al., 1960) were grown at 37°C in a humidified atmosphere of 5% CO<sub>2</sub> in Dulbecco's Modified Eagle Medium (Gibco) supplemented with 2mM glutamine, 100 IU of penicillin per ml, 100 IU of streptomycin per ml and either 10% lamb serum (growth medium) or 1% lamb serum (maintenance medium).

Transfections were performed by using ovine fetal synovial (FOS) cells. DNA was transfected with Lipofectamine as specified by the manufacturer (Life Sciences, Inc.). Transfected FOS cells were passaged and incubated in maintenance medium until syncytia appeared (5 to 8 days). Supernatants from transfected cells were also tested for the presence of reverse transcriptase (RT) activity before passage into SCP cells.

### **Generation of stable cell lines**

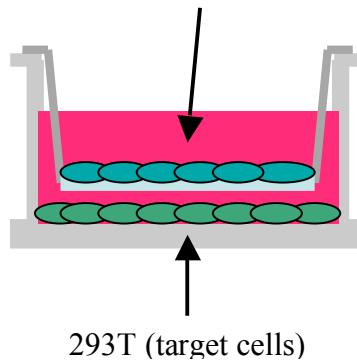
Stable APOBEC3G- or vector control-expressing PK-15 cell lines were constructed by transfection using FuGENE6 according to the manufacturer's protocol . Clones

were selected using growth medium containing 1 mg/ml G418 (Cordier et al., 1990), and APOBEC3G expressing clones were identified by immunoblotting using a polyclonal antibody toward human APOBEC3G. All PK-15 clones were maintained in growth medium supplemented with 250 µg/ml G418 to ensure stable expression.

### Long-term co-culture experiments

Long-term co-culture assays were performed in 6 well tissue culture plates with inserts (Transwell®, Corning Inc.). This system uses a membrane with 0.4 µm diameter pores, which keeps the two cell types separated physically but simultaneously allows diffusion of nutrients and small molecules including virus particles of approximately 0.1 µm (including PERV). Each experiment was initiated with 75,000 PK-15 cells (insert) and 75,000 293T cells (well) as illustrated (Figure 7). At 72 hr intervals, each cell type was washed with PBS, subjected to mild trypsinization and diluted into 4 parts fresh growth medium. Excess 293T cells were used to prepare genomic DNA (Qiagen DNeasy kit). The rate of PERV transfer was calculated using the *pol* gene levels from the last two data points (usually spanning a 3 day period). The difference between these levels represents PERV *pol* gene DNA that has accumulated per 100,000 human  $\beta$ -actin gene copies (50,000 cells assuming that the 293T cell line has two  $\beta$ -actin copies) per time period (Figure 8).

PK-15 (PERV producer cells (+/- A3))



293T (target cells)

**Figure 8.** Schematic of the co-culture system. PERV transmitting PK-15 are grown on top of the membrane of the insert and human 293T cells on the bottom of the well of the culture dish.

## DNA methods

### Artiodactyl A3F cDNA sequences and expression plasmids

*E. coli* expression plasmids were derived from pTrc99A (GE Healthcare). The SsA3F expression construct was made by PCR amplifying a cDNA [GB# BI346898] (Fahrenkrug et al., 2002) using 5'-NNN NGA GCT CAG GTA CCA CCA TGG ATC CTC AGC GCC TGA GAC and 5'- NNN NGT CGA CTC ATC TCG AGT CAC TTC TTG ATG, digesting the PCR product with KpnI and SalI and ligating it into a similarly digested pTrc99A. The BtA3F expression construct was made in the same manner by PCR amplifying a cDNA [GB# BE684372 (Smith et al., 2001)] using oligonucleotides 5'-NNN NGA GCT CAG GTA CCA CCA TGC AAC CAG CCT ACC GAG GC and 5'-NNN NGT CGA CCT AAA TTG GGG CCG TTA GGA T. OaA3F was obtained by first amplifying a fragment by degenerate PCR using oligonucleotides 5'-TWY RTV TCC TGG AGC CCC TG and 5'-CCR KWW GTW GTA GAG GCG R, which were made to conserved regions of APOBEC3 proteins using template cDNA from sheep macrophages. The amplified fragment was sequenced and used to make OaA3F-specific oligonucleotides for 3' prime RACE to obtain the remainder of the OaA3F coding sequence, 5'-AAC CAG GTC TAT GCT GGG ACT and 5'-CTG GGG ATG TAC CAG AAT GTG were used with an oligo dT primer 5'- AAG CAG TGG TAA CAA CGC AGA GTA CT30V N. The OaA3F expression plasmid was made by amplification from cDNA from sheep macrophages using oligonucleotides 5'-NNN NGA GCT CAG GTA CCA CCA TGC CCT GGA TCA GCG ACC AC and 5'-NNN NGT CGA CCT AAG TCG GCG CCG TCA GGA T in the same manner as the BtA3F and SsA3F constructs. E → Q zinc-binding domain mutants were constructed using standard site-directed mutagenesis protocols (Stratagene). The following oligonucleotides were used: SsA3F(E87 → Q) 5'-CCG ACC CGC CCT GCC ACG CCC AGC TCT GCT TCC TCT CTT GG and 5'-CCA AGA GAG GAA GCA GAG CTG GGC GTG GCA GGG CGG GTC GG; SsA3F(E267 → Q) 5'-ACA AGA AAA AGC GAC ATG CAC AAA TTC GTT TTA TTG ACA AG and 5'-CTT GTC AAT AAA ACG AAT TTG TGC ATG TCG CTT TTT CTT GT; BtA3F(E80 → Q) 5'-GTG GGA CTC GCT GCC ACA CCC AAC TCC GCT TCC TGT CTT GG and 5'-CCA AGA CAG GAA GCG GAG TTG GGT GTG GCA GCG AGT CCC AC; BtA3F(E260 → Q) 5'-ACA AGA AGC AGC

GGC ATG CAC AAA TTC GCT TTA TTG ACA AG and 5'-CTT GTC AAT AAA GCG AAT TTG TGC ATG CCG CTG CTT CTT GT; OaA3F(E67 → Q) 5'-CTG GGA CTC ACT GCC ACA GCC AAC GCC GCT TCC TGT CTT GG and 5'-CCA AGA CAG GAA GCG GCG TTG GCT GTG GCA GTG AGT CCC AG; OaA3F(E247 → Q) 5'-ACA AGA AGC AGC GGC ATG CAC AAA TTC GCT TTA TTG ACA AG and 5'-CTT GTC AAT AAA GCG AAT TTG TGC ATG CCG CTG CTT CTT GT. Eukaryotic expression plasmids were derived from pcDNA3.1(+) (Invitrogen) by subcloning the KpnI/SalI-flanked A3 cDNAs from the *E. coli* expression plasmids.

HsA3F, HsA3G and HsAID expression plasmids were described previously (Hache et al., 2005; Harris et al., 2002). MmA3 was made in the same manner as the artiodactyl constructs using oligonucleotides 5'-NNN NGA GCT CGG TAC CAC CAT GGG ACC ATT CTG TCT GGG A and 5'-NNN NGT CGA CAT CAA GAC ATC GGG GGT CCA AGC TG. E → Q zinc-binding domain mutants were constructed using standard site-directed mutagenesis protocols (Stratagene). The following oligonucleotides were used: HsA3G(E67 → Q) 5'-CTT AAG TAC CAC CCA CAG ATG AGA TTC TTC C and 5'-GGA AGA ATC TCA TCT GTG GGT GGT ACT TAA G; HsA3G(E259 → Q) 5'-CCT TGA AGG CCG CCA TGC ACA GCT GTG CTT CCT GG and 5'-CCA GGA AGC ACA GCT GTG CAT GGC GGC CTT CAA GG; HsA3F(E67 → Q) 5'-CTG AGC ACC ACG CAC AAA TGT GCT TCC TC and 5'-GAG GAA GCA CAT TTG TGC GTG GTG CTC AG; HsA3F(E251 → Q) 5'-CCC ATT GTC ATG CAC AAA GGT GCT TCC TC and 5'-GAG GAA GCA CCT TTG TGC ATG ACA ATG GG; MmA3(E73 → Q) 5'-CCA GTA TAA AAA GCA GAT TTG AGC GTG GAT GTT GTC CTT G and 5'-CAA GGA CAA CAT CCA CGC TCA AAT CTG CTT TTT ATA CTG G; MmA3(E290 → Q) 5'-CTT ATC AAG GAA GAG GAT TTG TGC ATG CTG TTT GCC TTT C and 5'-GAA AGG CAA ACA GCA TGC ACA AAT CCT CTT CCT TGA TAA G. Eukaryotic expression plasmids were derived from pcDNA3.1(+) (Invitrogen) and pEGFP-N3 (Clontech) by inserting KpnI and SalI or SacI and SalI flanked A3 cDNAs from the *E. coli* expression plasmids.

### ***A3Z1* gene degenerate PCR analyses**

Genomic DNA was isolated from the following tissues or cell lines: opossum kidney tissue, mouse NIH-3T3 cells, pig PK-15 cells, peccary brain tissue, cow MDBK cells, sheep FLK cells, horse blood cells (PBMC), African green monkey COS7 cells and human 293T cells (DNeasy, Qiagen). 10ng genomic DNA was used as template for PCR using primers designed to anneal to all known *A3Z1* genes: 5'- GCC ATG CRG AGC TSY RCT TCY TGG and 5'- GTC ATD ATK GWR AYT YKG GCC CCA GC-3'. Two PCR rounds were used to achieve the final number of cycles (30 plus 18, 21 or 24 cycles). Amplicons were analyzed by agarose gel electrophoresis, TOPO-cloned (Invitrogen) and subjected to DNA sequencing. In all instances, the expected *A3Z1* fragments were recovered (*e.g.*, Z1 of human *A3A*, *A3B* and *A3G* could all be detected in a single reaction). 30 PCR cycles using identical conditions and degenerate primers for the *ALDOA* gene were used as a positive control (5'-2 CGC TGT GCC CAG TAY AAG AAG GAY GG-3' and 5'-CTG CTG GCA RAT RCT GGC YTA).

### **Real-time PCR assay for MVV**

Viral particles from 200 µl of cell-free supernatants from infected cells were pelleted at 14,000 rpm for 1 hr in a microfuge. The pellet was dissolved in 10 µl TNE (10mM Tris pH 7.5; 100 mM NaCl; 1 mM EDTA) with 0.1% Triton X-100. This lysate was used for generating cDNA using RevertAid M-MuLV reverse transcriptase (Fermentas) and a primer from the *gag* gene (V-1818 5'CGG GGTACCTTACAACATAGGGGGCGCGG 3'). Real-time PCR was carried out with a DNA Engine Opticon System (MJ Research) in a final volume of 20 µl. The primers and Taqman probe were as follows: Forward primer: V1636 5' - TAAATCAAAAGTGTTATA ATTGTGGGA- 3', reverse primer: V-1719: 5' - TCCCACAATGATGGCATATTA TTC- 3', Taqman probe: V1665Taqman 5'-FAM-CCAGGACATCTCGCAAGA CAGTGTAACA-BHQ-1- 3'. Calibration curves were derived by running 10-fold dilutions of specific cDNA plasmid over the range of 60 to  $6 \times 10^7$  copies. Each assay included duplicate wells for each dilution of calibration plasmids and for each cDNA sample.

### **Genomic DNA Quantitative-PCR (qPCR) assays**

Genomic DNA was isolated from human 293T cells using the DNeasy kit (Qiagen). Duplicate 25 µl PCR reactions consisting of 10 ng of 293T genomic DNA, 100 nM primers and 2x iQ SYBR Green super mix (BioRad) were run on an iCycler iQ Multicolor Real-Time PCR detection System (BioRad). The thermocycler conditions consisted of an initial denaturation of 95°C for 5 min and 50 cycles of denaturation (95°C for 15 sec) and annealing (58°C for 30 sec). After the 50 cycles, a melting curve analysis (55°C to 95°C) was performed to confirm product specificity. The cycle threshold ( $C_T$ ) was generated using BioRad software and it was used to calculate the amount of target DNA (PERV *pol* or human  $\beta$ -actin). A standard curve was generated using the method of Dorak (Dorak, 2006) and a dilution series (10 to  $10^7$  copies) of a linearized plasmid containing the relevant 193 bp PERV *pol* gene fragment. The equation generated from the standard curve (slope and y intercept) was used to determine the efficiency of the PCR reaction and to quantify the number of PERV *pol* gene or human  $\beta$ -actin copies in the qPCR reactions. PERV copy numbers were normalized to those of  $\beta$ -actin. The primer sets used in this study were: PERV *pol* (193 bp): 5'-AAC CCT TTA CCC TTT ATG TGG AT and 5'-AAA GTC AAT TTG TCA GCG TCC TT; human  $\beta$ -actin (133 bp): 5'-ATC ATGTTT GAG ACC TTC AA and 5'-AGA TGG GCA CAG TGT GGG T; pig genomicDNA (341 bp product specific to intron 5 of the pig APOBEC3F locus): 5'-TGG GGA GTG TGG AAT TAA CG and 5'-GGG GGT TAA GAA CCC AAC AT

### **PERV *pol* Gene DNA sequence analyses**

Human 293T cell genomic DNA was prepared from terminal co-cultures and 50 ng was used for high fidelity, PERV *pol* gene-specific PCR reactions (Phusion polymerase; Finnzymes). 193 bp products were cloned using the Zero Blunt TOPO PCR Cloning kit (Invitrogen) and sequenced (University of Minnesota Advanced Genetic Analysis Facility). Sequence comparisons were performed using Sequencher software (Gene Codes Corp.) and publicly available Clustal W alignment algorithms (<http://align.genome.jp/>).

### RT-PCR experiments

Standard reverse transcription (RT)-PCR reactions were performed using RNA prepared from PK-15 cells (TRIzol protocol, Invitrogen), MMLV reverse transcriptase was used for cDNA synthesis using an oligo dT primer (Ambion) and Taq polymerase was used for PCR. The primers specific to pig APOBEC3F were 5'-TGG TCA CAG AGC TGA AGC AG and 5'-TTG TTT TGG AAG CAG CCT TT (175 bp). The semi-nested primer set used to detect plasmid expressed pig APOBEC3F was 5'-CCA AGG AGC TGG TTG ATT TC (exon 6, reaction 1), 5'-CTG GAG CAA TAC AGC GAG AG (exon 7, reaction 2) and 5'-TAG AAG GCA CAG TCG AGG, with the latter being vector specific (319 bp and 190 bp products, respectively). The mammalian  $\beta$ -actin primers were 5'-CCT TCA ATT CCA TCA TGA AGT G and 5'-CCA CAT CTG CTG GAA GGT (236 bp). These primers amplify equally well a 236 bp  $\beta$ -actin fragment from all mammals tested, including pigs and humans (e.g., Online Figure S3).

### APOBEC3 loci genomic DNA sequences

A combination of array hybridization, *A3*-, *CBX6*- and *CBX7*-specific PCR was used to identify one *A3*-positive BACs for sheep (CHORI-243 clone 268D23 <http://bacpac.chori.org/library.php?id=162>) and two for pigs (RPCI-44 clones 344O17 and 408D3). *E. coli* were transformed with these BACs, grown to saturation in 50 ml cultures and used for DNA preparations as recommended (Marligen Biosciences). Purified BAC DNA was sheared to an average of approximately 3000 bp (Hydroshear method, Genomic Solutions). Fragment ends were blunted with T4 and Klenow DNA polymerases (NEB) and ligated into pBluescriptSK- (Stratagene) or pSMART17 HC (Lucigen). Individual subclones were picked randomly and sequenced (ABI3730; Applied Biosystems). Phrap (P. Green, 1996, <http://www.phrap.org/phredphrap/phrap.html>) and Sequencher 4.8 (Gene Codes Corp.) were used to assemble DNA sequences and they were groomed manually. Sequence coverage for the sheep *A3* locus averaged 4.5 sequences and the pig 27 sequences. The genomic sequences were compared using Jdotter software (<http://www.jxxi.com/webstart/app/jdotter-a-java-dot-plot-viewer.jsp>; (Brodie et al., 2004)). Repetitive sequences were identified using RepeatMasker ([www.repeatmasker.org](http://www.repeatmasker.org)).

A3 exons were identified by directly comparing the genomic DNA sequences with cDNA, EST and RACE sequences. The sheep *CBX6* exons were identified with the help of GenBank EST sequences EE808826.1, DY519385.1 and EE822736.1. The pig *CBX6* exons were also identified in this manner using BP158234.1, BP997823.1 & BP153834.1. The sheep and pig *CBX7* exons were identified by homology to the cow gene. Other *CBX6* and *CBX7*, sequences, respectively, were NM\_014292.3 and NM\_175709.2 , NM\_001103094 and XM\_604126 (cow), NM\_028763.3 and NM\_144811 (mouse) and NM\_001016617.2 & NM\_001005071 (frog).

### **Amino acid alignments and phylogenetic analyses**

Protein alignments and phylogenetic analyses of the conserved, zinc-binding deaminase cores were done using the Clustal W software (Thompson et al., 1994). A rooted phylogenetic tree was constructed using the neighbor-joining method.

### **Infections and Infectivity assays**

#### **Experimental infection of sheep.**

Icelandic sheep, six months old were infected intratracheally with 1.0 ml of  $10^4$  TCID<sub>50</sub>. Virus was injected into the trachea with a needle (23G) under sedation with xylazine (Xylapan®, 0.2 ml I.V.) (Torsteinsdottir et al., 2003).

#### **Virus isolation from blood and organs**

Virus was isolated from buffy coat as described previously (Torsteinsdottir et al., 1997). The following organs were tested for presence of infectious virus by coculture with SCP cells (Petursson et al., 1976): choroid plexus, cerebrum, cerebellum, spleen, cervical, mediastinal and mesenteric lymph nodes, bone marrow and lungs (one sample from each lobe). If no cytopathic effects were observed, the growth medium was passaged twice, with the last observation after 6 weeks of culture.

#### **Green fluorescent protein (GFP)-based retrovirus infection assays**

293T cells were grown in DMEM supplemented with 10% FBS, penicillin and streptomycin. Viruses were produced by lipid-mediated transfection (Fugene; Roche)

of 50% confluent 293T cells with the following plasmids. HIV-GFP: 0.22 µg of CS-CG (Miyoshi et al., 1998), 0.14 µg of pRK5/Pack1(Gag-Pol), 0.07 µg of pRK5/Rev, 0.07 µg of pMDG (VSV-G Env) and 0.5 µg of an APOBEC3 expression plasmid or an empty vector control. In some experiments, an HIV-1 Vif or a ΔVif control plasmid was included (0.5 µg). MLV-GFP was produced similarly, except the proviral plasmid was pM3PGFP (0.2 µg) and a MLV Gag-Pol (0.14 µg) construct was used, along with 0.06 µg of pMDG (VSV-G Env) and 0.4 µg of an APOBEC3 expression plasmid or an empty vector control. After 48 hours of incubation, the virus-containing supernatants were clarified by low speed centrifugation, filtered (0.45 µm), and quantified using a reverse transcriptase activity-based ELISA (Cavidi Tech). Reverse transcriptase-normalized supernatants were applied to fresh 293T cells, and infection was allowed to proceed for 72 hr. Infectivity (GFP fluorescence) was measured by flow cytometry (FACSCalibur, BD Biosciences). For experiments requiring the recovery of retroviral DNA for hypermutation analyses, the viral supernatants were treated with 50 units/ml DNase (Sigma) prior to 293T cell infection.

### **Retroviral DNA sequence analyses**

Genomic DNA was incubated with DpnI to remove possible contaminating input CS-CGplasmid DNA. *GPF* was amplified using high fidelity Phusion polymerase (Finnzymes) and CSCG- specific primers, 5'-CG TGT ACG GTG GGA GGT CTA and 5'-TT GGT AGC TGC TGT GTT GCT. The PCR products were cloned directionally with internal NheI and XhoI sites into pBluescript (Stratagene) and sequenced using universal primers. Mutational analyses were performed using Sequencher software (Gene Codes Corp.).

### **Molecular assays**

#### **Reverse Transcriptase assay**

Reverse transcriptase (RT) activity of MVV was determined by taking 0.5 ml of cell-free supernatants from infected cells and pelleting them at 14,000 rpm and 4°C for 60 min in an Eppendorf 5417C microcentrifuge. The pelleted virus was resuspended in 10 ml of TNE buffer containing 0.1% Triton X-100. This disrupted virus was incubated at 37°C for 1 hour in a 50-ml reaction mixture containing 50mM Tris-HCl

(pH 7.8), 20mM MgCl<sub>2</sub>, 2mM dithiothreitol, 20mM KCl, 0.25 optical density unit each of poly (rA) and oligo (dT<sub>12</sub>±18) per ml, and 2.5 mCi of [3H]TTP (10±25 Ci/mmol; NEN). The reaction mixture (50 µl) was applied to DE81 filter paper discs (Whatman) and air dried. The filters were washed three times in 5% Na<sub>2</sub>HPO<sub>4</sub>, once in H<sub>2</sub>O, and then last in 70% ethanol. After drying, the filters were placed in vials with Ultima Gold (Packard) and radioactivity was measured using a Packard Tri-Carb liquid scintillation counter.

To determine RT activity of PERV whole cell protein extracts were prepared from 293T cells by suspending 500,000 cells in PBS, sonicating twice for 5 seconds and clarifying the lysates by centrifugation. Soluble protein levels were quantified using a BioRad Bradford assay. 10 µg of cell lysate was tested for reverse transcriptase activity using a C-type-RT activity assay (Cavidi Tech) following the manufacturers' instructions. Cell-free PK-15 supernatants (PERV-containing) were assayed directly using the Cavidi Tech ELISA assay

#### ***E. coli*-based deamination assay**

Intrinsic cytosine deaminase activity of the APOBECs is tested by expressing them in *E. coli* and monitoring mutation frequencies using the rifampicin-resistance (Rif<sup>R</sup>) mutation assay. Because *E. coli* naturally lacks an APOBEC-like DNA deaminase activity but possesses replicases capable of immortalizing the uracil lesion as a mutation, drugs such as rifampicin can be used to reveal changes in both mutation frequency and pattern caused by the expression of human cytosine deaminase proteins. Rifampicin binds the active site of the bacterium's RNA polymerase holoenzyme, thereby inhibiting its growth. Several amino acid changes affecting the active site confer drug resistance, and at the DNA level they map to a relatively small 200 bp interval in the bacterium's *rpoB* (RNA polymerase B) gene. Uracil DNA glycosylase-deficient *E. coli* strain BW310 was transformed with isopropyl 1-thio-β-D-galactopyranoside-inducible APOBEC expression constructs (above) or pTrc99A. Individual colonies were picked and grown to saturation in a rich medium containing 100 µg/ml ampicillin and 1 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG). Appropriate dilutions were plated onto a rich medium containing 100 µg/ml rifampicin or ampicillin to select Rif<sup>R</sup> colonies or measure cell viability, respectively, after an overnight incubation. Mutation frequencies were reported as the number of

Rif<sup>R</sup> colonies per 10<sup>7</sup> viable cells. *rpoB* sequences were obtained by colony PCR, Exo-SAP (USB) treatment of the PCR product, and direct DNA sequencing.

### **Fluorescent Microscopy**

One day prior to transfection, 20,000 cells were seeded onto LabTek chambered coverglasses (Nunc). After 24 hr incubation, these cells were transfected with 250 ng of the relevant plasmid construct. After an additional 24 hr of incubation, images of the live cells were acquired using a Zeiss Axiovert 200 microscope at 400x total magnification. Images were analyzed using Image J software (<http://rsb.info.nih.gov/ij>).

## 4. RESULTS

### **Simultaneous mutations in CA and Vif of Maedi-visna virus cause attenuated replication in macrophages and reduced infectivity in vivo (paper I)**

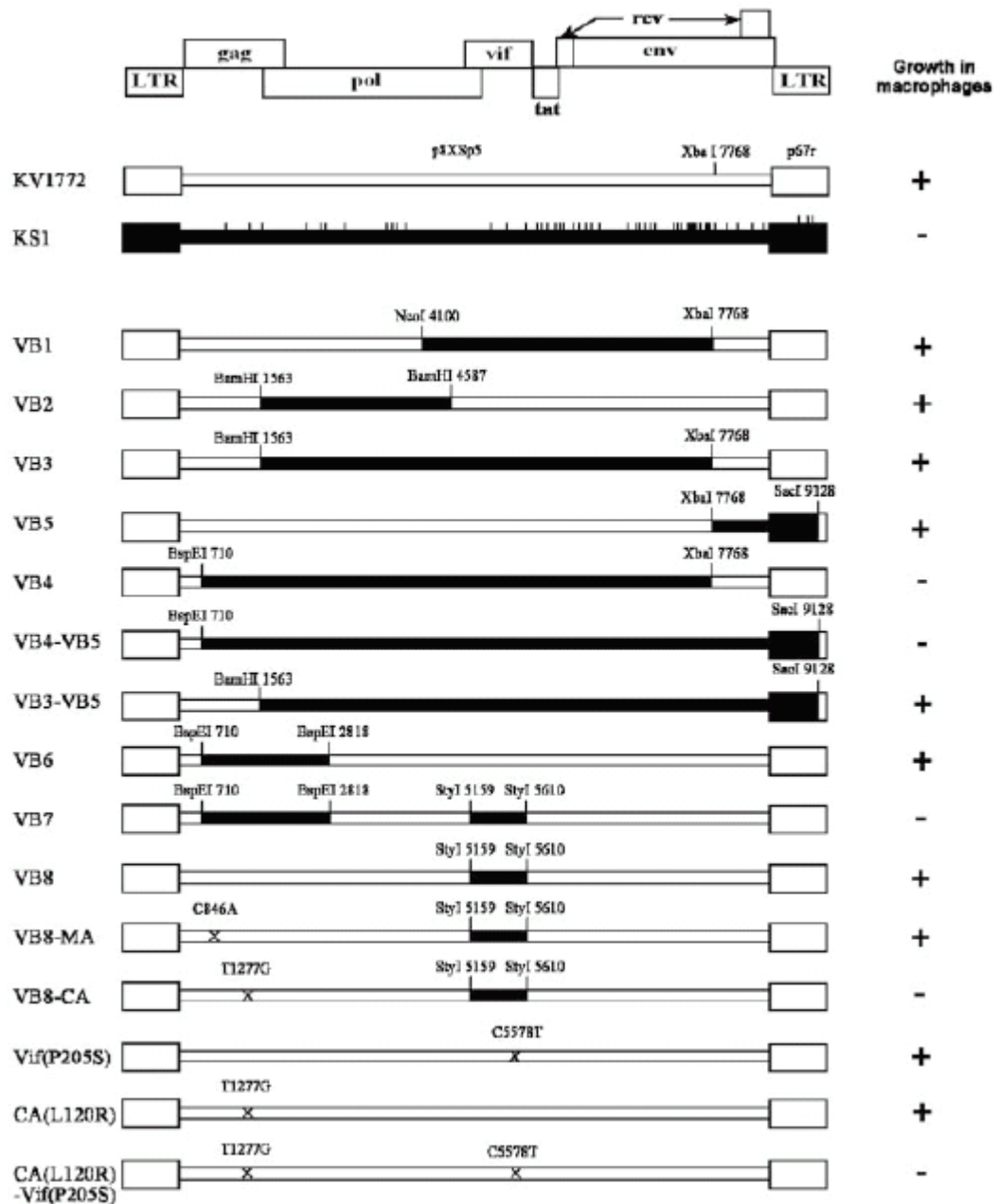
#### **The growth attenuation of the KS1 molecular clone is determined by regions of Vif and CA.**

In this study the different cell tropism and pathogenicity of two molecular clones of MVV was examined (Andresson et al., 1993; Staskus et al., 1991). One of the clones (KS1) replicates well in sheep choroid plexus (SCP) cells, but replicates only to low titers in macrophages and is non- pathogenic in sheep. The other clone (KV1772) replicates to high titers in SCP cells, FOS cells and macrophages and is highly pathogenic in sheep. Both clones have their origin in a transmission experiment where virus was passaged serially through tissue culture and sheep and differ only by 1% in nucleotide sequence. By constructing a series of recombinant viruses the attenuated phenotype of KS1 was mapped to two point mutations, one in the CA domain of *gag* and the other in *vif* (Figure 9). Interestingly, the P-S mutation in Vif is in a proline-rich region at the C-terminal end of the protein that has been shown to be important for association with cellular membranes and interaction with the Gag precursor Pr55<sup>gag</sup> (Bouyac et al., 1997a; Goncalves et al., 1994; Goncalves et al., 1995) whereas the mutation in CA is an L-R mutation in helix 7.

#### **In vivo inoculation with the two parental strains and the recombinant virus CA(L120R)-Vif(P205S).**

The KS1 strain has repeatedly been shown to be non-pathogenic in sheep, whereas the KV1772 strain is highly pathogenic (Torsteinsdottir et al., 1997). Of 219 attempts to isolate virus from blood of sheep infected with KS1 in various experiments, virus was isolated only on one occasion. Sequencing of the CA and *vif* genes of this isolate revealed a reversion of the L-R mutation in CA back to L. This isolate grew well in macrophages.

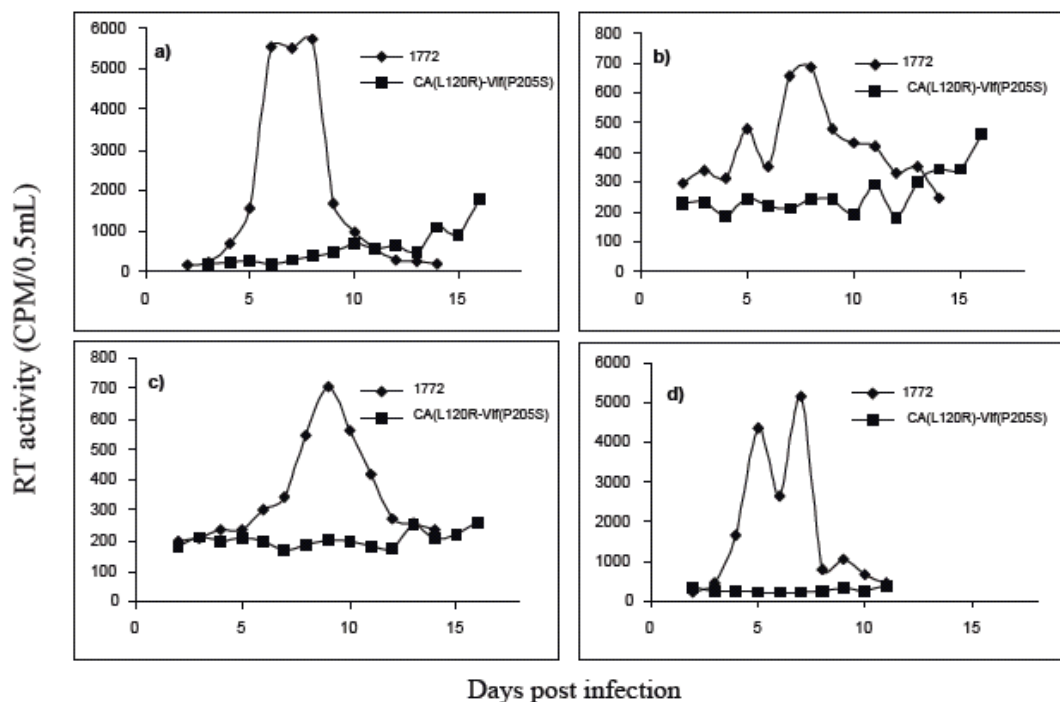
Three sheep, 6 months old, were inoculated intratracheally with each of the three virus strains, KV1772, KS1 and CA(L120R)-Vif(P205S). Blood was collected for virus isolation once a week for the first 8 weeks and biweekly thereafter until sacrifice after 18 weeks. At sacrifice virus isolations was attempted from several



**Figure 9.** Diagrammatic representation of recombinant MVV derived from the molecular clones KV1772 and KS1. The vertical bars indicate the amino acid differences between the two clones. The restriction sites used to construct the viruses are shown. The ability of each virus to grow in macrophages is shown on the right (Figure courtesy of Bjarki Guðmundsson).

organs, including brain, lungs, spleen, lymph nodes and bone marrow. As shown in Table 3, virus was readily recovered from blood and organs of the three sheep that had been inoculated with strain KV1772, whereas no virus was isolated from sheep inoculated with KS1 virus. However, one of the three sheep inoculated with the recombinant CA(L120R)-Vif(P205S) virus was clearly infected. This result was not

expected, since the CA(L120R)-Vif(P205S) strain has similar replication kinetics in macrophages as the KS1 strain. Sequencing of the virus isolates did not show a reversion of the mutations in *CA* and *vif*. One of the blood isolates was tested for replication in macrophages and it showed attenuated replication similar to the CA(L120R)-Vif(P205S) strain (data not shown). We then speculated that this particular sheep might not express a putative virus inhibitor. We therefore infected four additional sheep with the virus CA(L120R)-Vif(P205S), and isolated macrophages from each sheep before infection for testing replication of the CA(L120R)-Vif(P205S) strain compared to KV1772. Replication of the CA(L120R)-Vif(P205S) strain was attenuated in macrophages from all of the four sheep (Figure 10). The sheep were kept for 18 weeks as before, and attempts were made to isolate virus from the same organs. No virus was isolated from any of the organs, but one isolate was recovered from blood (Table 3).



**Figure 10.** Growth of viruses in macrophages from 4 different sheep. Macrophages from 4 sheep were isolated and infected with 1772 and CA(L120R)-Vif(P205S) and their growth measured by RT activity. A) sheep 2089. B) sheep 2090. C) sheep 2091 and D) sheep 2092.

Virus strain and sheep no.	No. positive/no. tested in:			
	Blood	Lymph <sup>a</sup>	Lungs	Central nervous system
KV1772				
2080	10/11	5/5	2/2	1/3
2081	8/11	5/5	2/2	0/3
2082	5/11	2/5	0/2	0/3
KS1				
2074	0/11	0/5	0/2	0/3
2075	0/11	0/5	0/2	0/3
2076	0/11	0/5	0/2	0/3
CA(L120R)-Vif(P205S)				
2077	0/11	0/5	0/2	0/3
2078	2/11	4/5	2/2	0/3
2079	0/11	0/5	0/2	0/3
2089	0/11	0/5	0/2	0/3
2090	0/11	0/5	0/2	0/3
2091	1/11	0/5	0/2	0/3
2092	0/11	0/5	0/2	0/3

<sup>a</sup> Spleen, bone marrow, cervical, mediastinal, and mesenteric lymph nodes.

**Table 3.** Frequency of virus isolation from KV1772-, KS1-, and CA(L120R)-Vif(P205S)-infected sheep

### Evolutionary conserved and non-conserved retrovirus restriction activities of artiodactyl double-domain APOBEC3 (APOBEC3F) proteins (paper II)

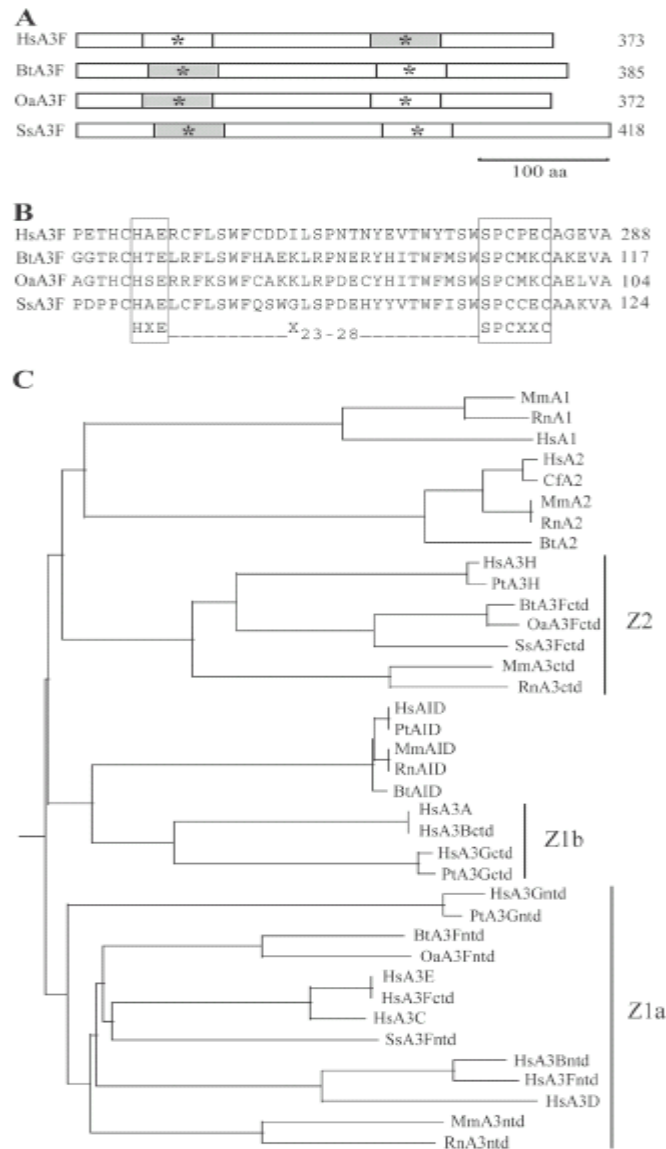
The initial aim of these studies was to clone and characterize the APOBEC3 proteins of sheep, due to the fact that little sheep DNA sequence exists in public databases. Therefore the project was broadened to include other closely related artiodactyl species. These studies focused on APOBEC3 proteins from three representative artiodactyl species: Cow (*Bos taurus*), sheep (*Ovis aries*) and pigs (*Sus scrofa*), that were predicted to be of similar size and domain organization as the well characterized human A3F and A3G, the most potent anti-retroviral human A3 proteins, as well as mouse A3.

### **The double deaminase domain APOBEC3 proteins of artiodactyls**

All known A3 proteins have either one or two conserved, zinc-binding deaminase domains, consisting of amino acids HXE-X<sub>23-28</sub>-PCX<sub>2-4</sub>C (X can be nearly any amino acid). The histidine and the two cysteines coordinate zinc and the glutamate participates directly in the C → U deamination reaction. NCBI BLAST searches using the human and mouse A3 deaminase domains as query polypeptides revealed several artiodactyl ESTs, which suggested the presence of at least one A3 protein in cattle and pigs. Corresponding cDNA clones were obtained, sequenced and shown to encode A3 proteins with two putative zinc-binding, cytosine deaminase domains (Figure 11). The orthologous sheep double domain A3 cDNA sequence was obtained using a combination of degenerate PCR and RACE (Figure 11). All three of these A3 proteins were similar in size to the double-domain human A3 proteins, except the pig A3 protein, which was slightly longer due to a unique C-terminal, serine-rich extension (Figure 11).

Nomenclature standards dictate that protein names should be assigned based on the closest human orthologue. However, amino acid comparisons showed that the N- and the C-terminal artiodactyl deaminase domains were most similar to different human A3 proteins. To resolve this ambiguity and to facilitate name assignments, we named the artiodactyl double domain proteins after ‘the human double domain A3 protein with the highest degree of active site identity’. The active sites of these artiodactyl A3 proteins were 56-62% identical to HsA3F (Figure11). Thus, the double domain deaminase proteins from cow, sheep and pig were named after HsA3F and, were referred to as BtA3F, OaA3F and SsA3F, respectively in the paper. Amino acid alignments of the active deaminase domains plus five residues on each side showed that the cow and sheep A3F active sites are 78% identical (Figure11). Both the cow and the sheep proteins shared a lower level of identity with the pig protein (56%). Since then the naming system for non-primate A3 proteins has been revised and that naming system will be used in the thesis.

As described above, all A3 proteins have either one or two conserved, zinc-binding (Z) deaminase domains. These domains cluster into three distinct phylogenetic groups: Z1, Z2 or Z3. The human double Z domain proteins, HsA3F and HsA3G, have a Z2/Z2 and a Z2/Z1 organization, respectively, whereas the MmA3 protein has a Z2/Z3 organization. Interestingly, all three of the artiodactyl A3F proteins have a Z2/Z3 organization (Figure11).



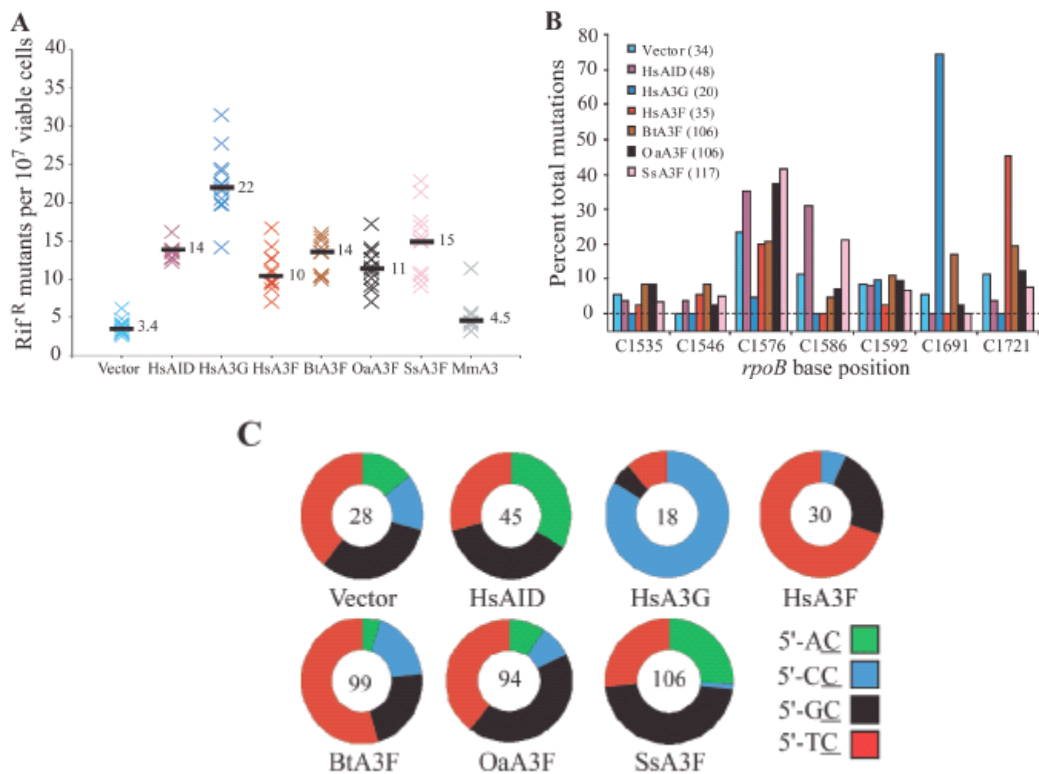
**Figure 11.** A comparison of artiodactyl A3F proteins. (A) A schematic of HsA3F, BtA3F, OaA3F and SsA3F. The conserved, zinc-binding deaminase domains are boxed (\*) and those that are catalytically active are additionally shaded. The numbers on the right indicate the total polypeptide length. (B) HsA3F, BtA3F, OaA3F and SsA3F active site amino acid alignments [shaded regions from (A)]. The conserved HXE and PCXXC motifs are boxed. Amino acid positions are indicated on the right. (C) A neighbor-joining phylogenetic tree indicating the evolutionary relationship of several representative mammalian A3 family members. Branch lengths are proportional to the number of amino acid differences. Comparisons were done using the conserved Z domain amino acids, plus five additional residues on either side. The Z1a, Z1b and Z2 phylogenetic clusters are indicated. HsA3D and HsA3E represent the N- and C-terminal domains of HsA3DE. See the text for additional details.

### **The artiodactyl double domain A3 proteins catalyze DNA cytosine deamination**

To test whether the artiodactyl A3Z2-Z3 proteins have the capacity to deaminate cytosines within ssDNA, the intrinsic mutator activity of these proteins was monitored using an *E. coli*-based mutation assay. Rifampicin resistance (Rif<sup>R</sup>) is attributable to base substitution mutations in the *E. coli* RNA polymerase B (*rpoB*) gene, and it occurs in approximately one of every five million bacterial cells. Expression of several A3 family members can accelerate the accumulation of Rif<sup>R</sup> mutations from a few- to several hundred-fold. The mutator phenotype is accounted for by a pronounced C/G → T/A transition bias within *rpoB*. This assay therefore provides a robust measure of intrinsic DNA cytosine deaminase activity.

Expression of each of the artiodactyl A3 proteins increased the Rif<sup>R</sup> mutation frequency in *E. coli* from 3- to 7-fold, levels that were higher than those attributable to HsA3F but slightly lower than those caused by HsA3G (Figure 12). Curiously, expression of MmA3 failed to cause an *E. coli* mutator phenotype, despite the fact that it is clearly active and capable of deaminating the cDNA of a variety of retroelements. It is not clear why MmA3 is inactive in this system, yet active in others.

Artiodactyl A3F DNA cytosine deamination preferences were examined by sequencing the *rpoB* gene of at least 100 independent Rif<sup>R</sup> mutants (Figure 12B and C). In contrast to HsA3F and HsA3G, which preferentially deaminate cytosines at *rpoB* nucleotide positions 1721 and 1691, 5'-TC and 5'-CC, respectively, the artiodactyl A3 proteins showed less biased *rpoB* mutation spectra. Sheep A3Z2-Z3 preferentially deaminated cytosine 1576, which is part of a 5'-GC dinucleotide. Pig A3Z2-Z3 also preferred cytosine 1576. However, the pig protein also clearly deaminated cytosine 1586, which is part of a 5'-AC dinucleotide. Interestingly, these two cytosines, C<sub>1576</sub> and C<sub>1586</sub>, are also preferred by HsAID (Figure 12B). The cow A3Z2-Z3 did not appear to have any prominent *rpoB* local mutation preference, as increased levels of C/G → T/A mutation were apparent at several sites. In conclusion, all three of the artiodactyl A3F proteins are capable of deaminating DNA cytosines to uracils, which triggers a corresponding shift in the pattern of C/G → T/A transition mutations within the *rpoB* mutation substrate.

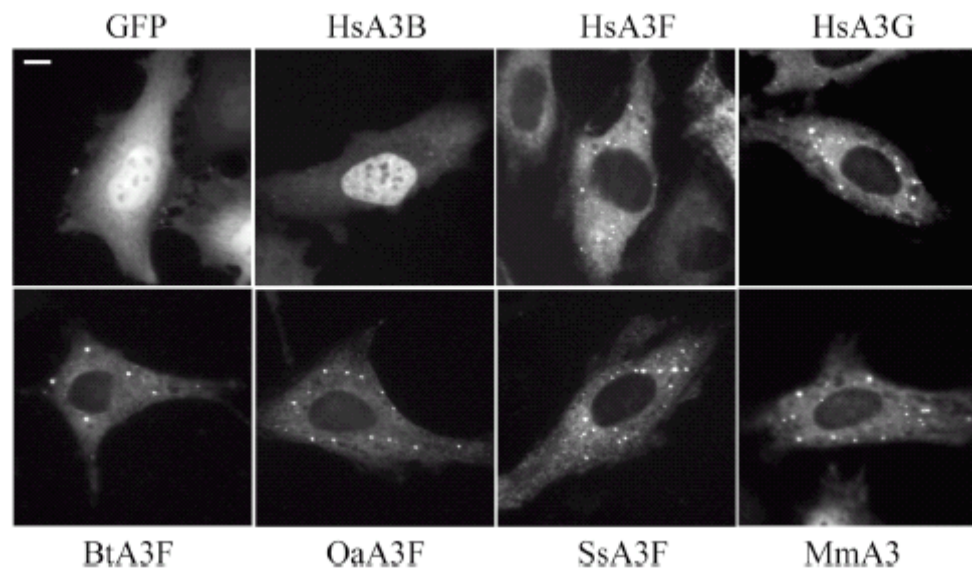


**Figure 12.** DNA cytosine deaminase activity of the artiodactyl A3F proteins in *E. coli*. (A) Rif<sup>R</sup> mutation frequencies for 12 independent *E. coli* cultures expressing a vector control (light blue), HsAID (purple), HsA3G (dark blue), HsA3F (red), BtA3B (brown), OaA3B (black), SsA3F (pink) or MmA3. Each data point corresponds to the mutation frequency obtained from a single culture, and the median mutation frequency for each condition is shown (horizontal bar). (B) A histogram summarizing the C/G→T/A transition mutations detected in *rpoB*. Only cytosines that had greater than two mutations are shown. Apart from C1586, all of the cytosines are located in the non-template strand of the *rpoB* gene. The number of independent Rif<sup>R</sup> mutants that were sequenced is indicated in parentheses in the legend. For purposes of presentation the Y-axis extends below zero (a dotted line marks the actual base-line), and the histogram bars follow the color scheme (A). (C) Pie graphs depicting the frequency that each of the four dinucleotides was targeted by the indicated A3 protein. The total number of independent sequences analyzed is shown in the center of each graph, and the deaminated cytosine is underlined in the legend. The dinucleotide wedges of each pie are colored as indicated.

### The artiodactyl double domain A3 proteins are predominantly cytoplasmic

As an initial step toward understanding the potential retroelement targets of the artiodactyl A3F proteins, the sub-cellular distribution of these proteins was determined. A3-GFP constructs were transfected into HeLa cells and the sub-cellular

localization of the A3 proteins was determined by live cell fluorescence microscopy. In agreement with prior work, HsA3B and a GFP-only control localized to the nucleus and cell-wide, respectively. In contrast, the artiodactyl A3Z2-Z3 proteins and MmA3 appeared predominantly cytoplasmic (Figure 13). Several of the images also revealed brightly fluorescing, punctate cytoplasmic aggregations, which may represent P bodies. The significance of the cytoplasmic punctae remains to be determined. Nevertheless, the cytoplasmic localization pattern is nearly identical to that of HsA3F and HsA3G suggesting that the property of localizing to the cytoplasm is conserved and that the artiodactyl A3Z2-Z3 proteins might function similarly to inhibit the replication of LTR-dependent retroviruses such as HIV-1 or MLV (Figure 13).



**Figure 13.** Sub-cellular distribution of the artiodactyl A3Z2-Z3 proteins in comparison to the orthologous human and mouse A3 proteins. HeLa cells showing localization of the indicated, GFP-tagged A3 proteins or a GFP-only control. The scale bar indicates 10 mm (figure courtesy of Mark D. Stenglein).

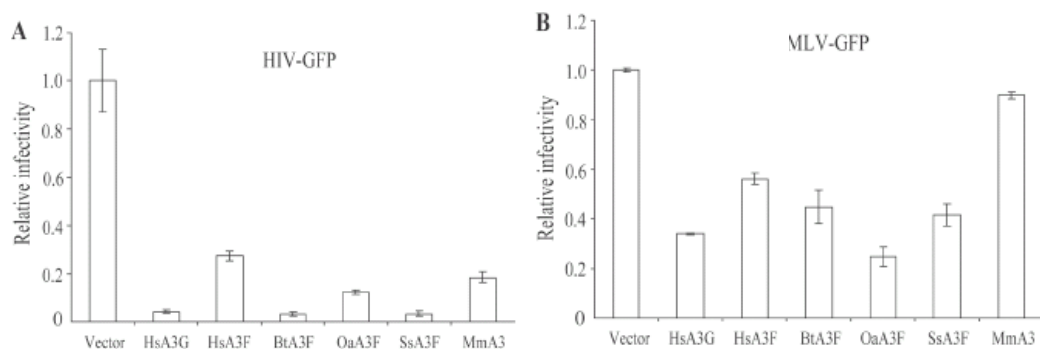
### Retrovirus restriction by double domain artiodactyl A3 proteins

A clear trend in the genetic conflict between A3 proteins and retroelements is that an A3 from a given host is either neutralized or avoided by retroelements that are specific to the host species. In many instances, cross-species comparisons enable potential species-specific mechanisms of neutralization to be avoided and the restrictive potential of A3 proteins to be studied. We therefore asked whether the

artiodactyl A3Z2-Z3 proteins could inhibit the infectivity of HIV-1 and MLV-based retroviruses. In these systems, a *GFP* gene embedded in proviral DNA provides a measure of both transfection efficiency (which correlates directly with virus production levels) and viral infectivity.

Expression of HsA3F and HsA3G caused 4- and 24-fold reductions in the infectivity of HIV-GFP (Figure 14A). MmA3 was also capable of strongly inhibiting HIV-GFP. In comparison, expression of BtA3Z2-Z3, OaA3Z2-Z3 or SsA3Z2-Z3 caused 30-, 8- and 29-fold decreases in the infectivity of HIV-GFP, respectively (Figure 14A). These potent anti-HIV activities demonstrated that the artiodactyl A3F proteins possess at least one retrovirus restriction activity. These results further imply that the artiodactyl A3Z2-Z3 proteins are able to specifically associate with the HIV-1 Gag/genomic RNA complex and thereby gain access to assembling virus particles.

Expression of MmA3 has little effect on the infectivity of MLV, presumably because MLV excludes (or simply avoids) this A3 protein (Figure 14B). In contrast, HsA3F and HsA3G inhibit the infectivity of MLV-based retroviruses, but to a lesser extent than HIV lacking Vif (Figure 14B). Therefore, to ask whether the artiodactyl A3Z2-Z3 proteins possess similar restriction potentials, the infectivity of MLV-GFP produced in the presence of these A3 proteins was monitored. Interestingly, similar to HsA3F or HsA3G, expression of the artiodactyl A3Z2-Z3 proteins reduced the infectivity of MLV-GFP by 2- to 4-fold. Thus, the HIV-GFP and MLV-GFP infectivity data combined to suggest that the artiodactyl A3Z2-Z3 proteins have a relatively broad retrovirus restriction potential.



**Figure 14.** Retrovirus restriction activity of the artiodactyl A3Z2-Z3 proteins.

(A) Infectivity of HIV-GFP produced in the presence of a vector control or the indicated A3 protein. Data were normalized to the infectivity of HIV-GFP produced in the presence of a control vector, which was assigned a value of one. The mean and the SEM of three independent experiments are shown.

(B) Infectivity of MLV-GFP produced in the presence of the indicated constructs. Parameters are identical to those in (A).

**The N-terminal zinc-binding, deaminase domain of the artiodactyl double domain A3 proteins catalyzes C → U deamination, and this activity is necessary for full levels of retrovirus restriction**

All of the double domain deaminases thus far characterized have catalytically competent C-terminal Z domains and apparently inert N-terminal Z domains. Most experiments have focused on HsA3B, HsA3F and HsA3G, DNA cytosine deaminases that have Z1- or Z2-type, C-terminal active sites (Figure 11). However, HsA3H, which has a single Z3-type deaminase motif, was also shown to possess DNA cytosine deaminase activity. Thus, it was possible that either the N- or the C-terminal Z domain (or both) of the artiodactyl A3Z2-Z3 proteins would be catalytically active.

To begin to work out the mechanism of retrovirus restriction by artiodactyl A3Z2-Z3 proteins and to test whether the N- or the C-terminal (or both) Z domains of these proteins catalyze DNA cytosine deamination, the conserved glutamate (E) of each active site was changed to glutamine (Q) and the resulting mutants were tested for HIV-GFP restriction activity. As reported previously, the glutamate of both the N- and the C-terminal Z domain of HsA3G contributed to inhibiting HIV-1 infectivity, but the C-terminal catalytic glutamate appeared to be more important (Figure 15A). In contrast, both the N- and the C-terminal BtA3Z2-Z3 Z domain E → Q mutants appeared to retain full levels of anti-HIV activity. Interestingly, the N-terminal OaA3Z2-Z3 and SsA3Z2-Z3 Z domain E → Q mutants were less able than the corresponding C-terminal domain mutants to inhibit the infectivity of HIV-GFP, a result particularly clear for SsA3F (Figure 15A). These data were essentially the inverse of the HsA3F and HsA3G E → Q mutant studies, and they suggested that the N-terminal, Z domain of these proteins catalyzes retroviral cDNA C → U deamination. MmA3 was clearly distinct, as both the N- and the C-terminal Z domain glutamates were required for HIV-GFP restriction (Figure 15A).

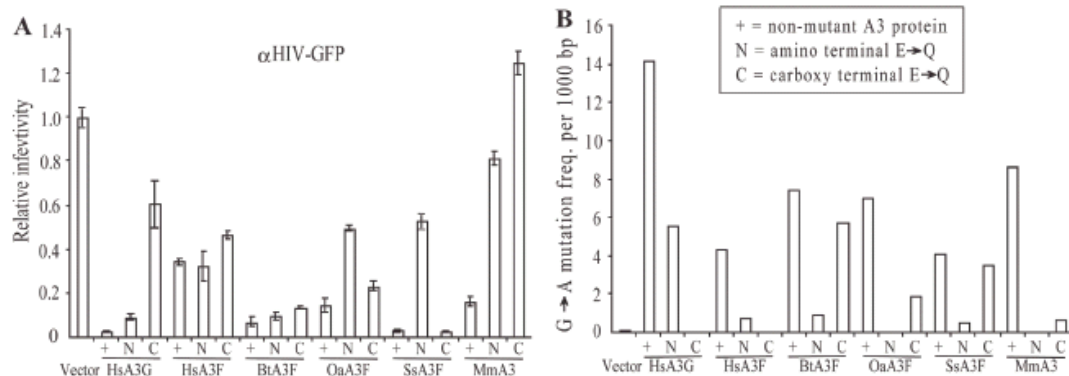
Although both the N- and the C-terminal Z domain E → Q mutants of the human and the artiodactyl A3 proteins showed significant levels of anti-retroviral activity, we surmised that *bona fide* catalytic site mutants should be unable to catalyze retroviral cDNA C → U deamination (although they may still inhibit retroviral infectivity). Minus strand uracils template the incorporation of plus strand adenines, ultimately manifesting as retroviral plus strand G → A hypermutations. Therefore, to directly test which Z domain(s) catalyzes DNA cytosine deamination

and to gain additional insight into the artiodactyl A3Z2-Z3 retrovirus restriction mechanism, the *GFP* gene from the aforementioned HIV-GFP infectivity experiments was amplified by high-fidelity PCR, cloned and sequenced. HIV-GFP produced in the presence of a control vector showed a low base substitution mutation frequency, 0.00014 mutations per base, which is attributable to errors in reverse transcription and PCR (Figure 15B). In contrast, viruses produced in the presence of HsA3F, HsA3G, all three of the artiodactyl A3Z2-Z3 proteins or MmA3 showed a 30- to 80-fold increase in base substitution mutations, which were almost exclusively plus-strand G → A transition mutations (Figure 15B). HsA3G with a C-terminal domain E → Q mutation failed to cause retroviral hypermutation, although this variant still significantly inhibited HIV-GFP infectivity (Figure 15). The HsA3F C-terminal Z domain mutant was still able to modestly inhibit HIV-GFP infectivity, without obvious signs of retroviral hypermutation. Interestingly, E → Q substitutions in the N-terminal (but not the C-terminal) domain of all three of the artiodactyl A3Z2-Z3 proteins abolished the accumulation of retroviral hypermutations (Figure 15B). Thus, these data combined to demonstrate that the N-terminal Z domain of the artiodactyl A3F proteins is catalytic and that both deaminase-dependent and -independent activities are required for full levels of retrovirus restriction. In support of this conclusion, the N-terminal Z domain glutamate of BtA3Z2-Z3 is required for mutator activity in *E. coli*, whereas the C-terminal Z domain glutamate is dispensable.

### **Retroviral hypermutation properties of artiodactyl double domain A3 proteins**

As described above, the *rpoB* mutation spectra of BtA3Z2-Z3, OaA3Z2-Z3 and SsA3Z2-Z3 suggested that these proteins would trigger retroviral hypermutation patterns biased toward 5'-YC, 5'-GC, and 5'-RC, respectively (R = A or G; Y = C or T; Figure 12C). To test this prediction, we examined the types of base substitution mutations and the local retroviral cDNA deamination preferences in HIV-1 attributable to expression of the artiodactyl A3F proteins (Figure 15B). In terms of the dinucleotide mutation preferences, the base immediately 5' of the targeted cytosine is a crucial target site determinant. HsA3F and HsA3G overwhelmingly preferred 5'-TC (84%) and 5'-CC (84%), respectively, whereas MmA3 preferred 5'-TC (61%) and 5'-CC (29%). Roughly paralleling the *E. coli* Rif<sup>R</sup> mutation data,

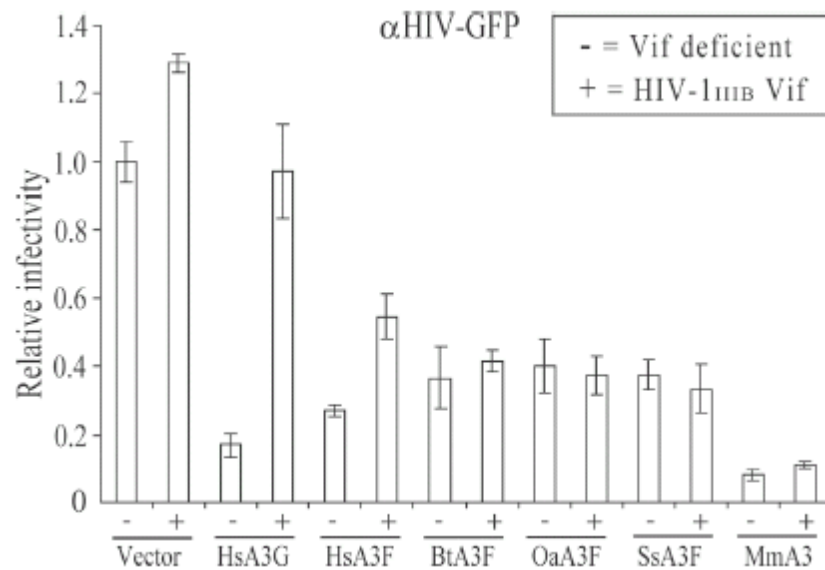
HIV-GFP sequences revealed that the cow and the sheep A3Z2-Z3 proteins preferred a pyrimidine (Y) 5' of the deaminated cytosine (93% and 79%, respectively). Similarly, the pig A3Z2-Z3 protein preferred 5'-GC (47%). This is notable because this is one of very few A3 proteins preferring 5'-purine-C (the proposed ancestor of the modern day A3 proteins, the immunoglobulin gene deaminase AID also has this preference). In addition, all of the A3 proteins characterized preferred a pyrimidine at the -2 position, T in all cases except for HsA3G which preferred C.



**Figure 15.** Relative contributions of the N- and C-terminal zinc-binding domains to HIV-GFP restriction. (A) Infectivity of HIV-GFP produced in the presence of a vector control or the indicated non-mutant (+) or mutant A3 protein containing an N-terminal E!Q substitution or a C-terminal E → Q substitution. (B) Frequencies of plus-strand retroviral G → A hypermutation observed in HIV-GFP DNA.

### The artiodactyl double domain A3 proteins are fully resistant to HIV-1 Vif

HsA3F, HsA3G and chimpanzee A3G are neutralized by HIV-1 Vif. However, many other monkey A3G proteins (*e.g.*, African green monkey) and MmA3 are completely resistant. The full sets of interactions that govern the A3-Vif conflict have not been determined, and the artiodactyl A3F proteins are likely to prove useful in this regard. Therefore, HIV-GFP infectivity was monitored in the presence or absence of HIV-1 Vif and human, artiodactyl or mouse A3 proteins. As described previously, expression of HIV-1 Vif neutralized HsA3F and HsA3G (although the former to a lesser extent) and caused a proportional recovery of HIV-GFP infectivity (Figure 16). Expression of HIV-1 Vif failed to enhance the infectivity of HIV-GFP produced in the presence of MmA3 or any of the artiodactyl A3Z2-Z3 proteins. Thus, the artiodactyl A3Z2-Z3 proteins were fully resistant to HIV-1 Vif.



**Figure 16.** Artiodactyl A3Z2-Z3 proteins are resistant to HIV-1 Vif. Parameters are identical to those used in Figure 4A, except for the inclusion of plasmids encoding HIV-1IIIIB Vif (+) or a HIV-1IIIIB DVif (-) control, which has translation stop codons at amino acids 33 and 34.

### **The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals (paper III)**

The fact that mammals have varying numbers of A3 genes, with humans having seven and mice only having one, led us to wonder whether there were additional A3 genes in the artiodactyls to those we had worked with in the context of paper II. To address this point and to learn more about the evolution and functionality of A3 genes in mammals, we sequenced and characterized the full A3 repertoire of sheep and pigs and compared to available information from the cow.

#### **Sheep and cattle have three A3 genes with a Z1-Z2-Z3 organization**

We previously used degenerate PCR, RACE and database mining to identify a cDNA for sheep A3Z2-Z. To address the number of A3 genes in artiodactyls unambiguously, we sequenced the entire sheep A3 genomic locus. First, a sheep A3Z2-Z3 cDNA was hybridized to a sheep BAC library to identify a corresponding genomic sequence. Second, hybridization-positive BACS were screened by PCR for

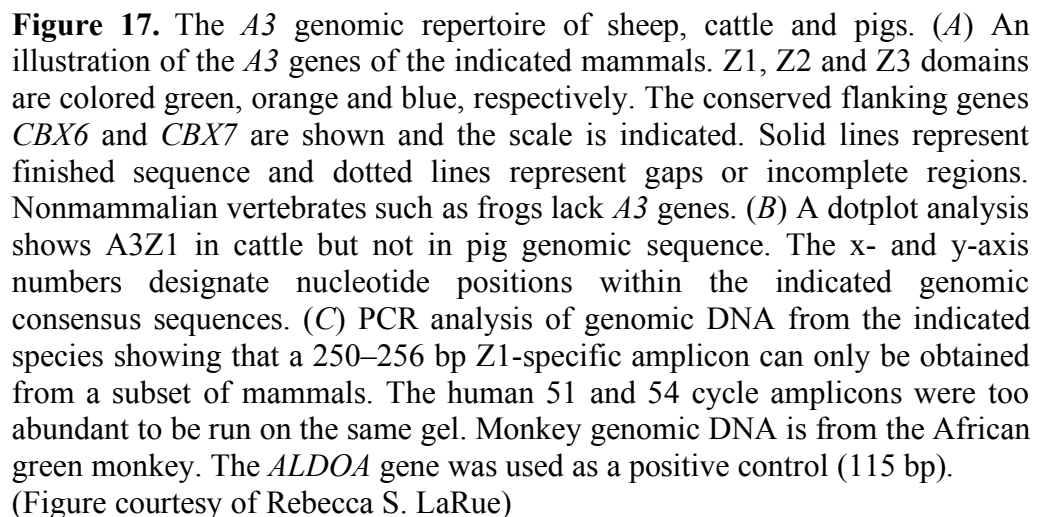
those that also contain the conserved flanking genes *CBX6* and *CBX7*. One BAC was identified that spanned the entire *CBX6* to *CBX7* region, and it was sheared, subcloned, shotgun sequenced, assembled and analyzed. DNA sequence analyses revealed that the sheep genomic locus contained an additional *A3* gene between *CBX6* and *A3Z2* (Figure 17A). This gene was called *A3Z1*, because it had sequence characteristics of a Z1-type A3 protein. We therefore concluded that sheep have three *A3* genes and, importantly, that each mammalian *A3* Z-type was present. This conclusion was supported by the bovine genome assembly, which was released during the course of our studies and showed that cattle also have a sheep-like, three gene *A3* repertoire (Figure 17A). The predicted *A3Z1* coding sequences of sheep and cattle are 86% identical, consistent with the fact that these two ruminant artiodactyls shared a common ancestor approximately 14-25 million years ago (Bininda-Emonds et al., 2007).

#### **The pig has two *A3* genes with a Z2-Z3 organization**

PCR reactions failed to identify an *A3Z1*-like gene in pigs. Since pigs and cattle/sheep last shared a common ancestor approximately 70-80 MYA, we considered the possibility that the negative PCR result was not a technical failure and that pigs might actually have a different *A3* repertoire. Again, to unambiguously address this possibility, the pig *A3* genomic locus was sequenced in entirety. A porcine BAC library was probed with pig *A3Z2-Z3* cDNA and two hybridization-positive BACS were shotgun sequenced. The sequence assemblies revealed that pigs have only two *A3* genes, *A3Z2* and *A3Z3*, between *CBX6* and *CBX7* (Figure 17A).

The cattle, sheep and pig *A3* locus genomic sequences were compared using dotplot analyses (Figures 17B). A 22 kb discontinuity was detected between the cow and the pig sequences. The sheep and pig genomic sequences aligned similarly.

To begin to address whether the potential *A3Z1* deletion in pigs occurred recently (*e.g.*, a rare deletion fixed by selective breeding) or whether it was more ancient, we asked whether a non-domesticated, distant relative of the pig, the collared peccary (*Tayassu tajacu*), has an *A3Z1* gene. Lineages leading to present-day domesticated pigs and the peccary diverged approximately 25-35 MYA. A pan-species, *A3Z1* PCR primer set was developed and used in these experiments. In contrast to human, African green monkey, horse, cow and sheep genomic DNA which yielded a 250-



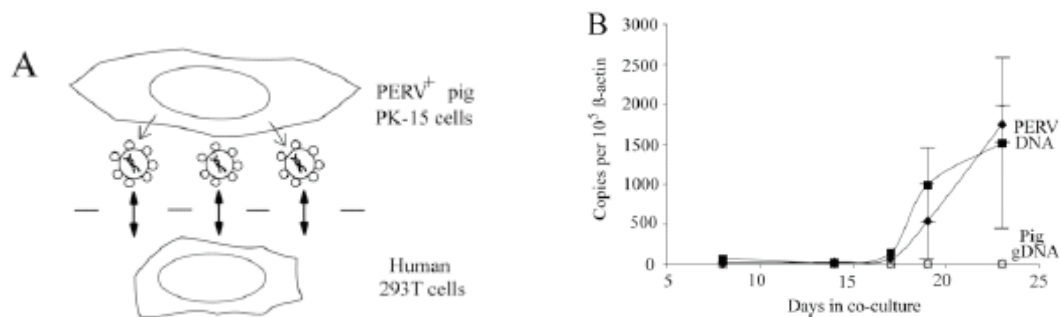
## **The restriction of Zoonotic PERV transmission by human APOBEC3G (paper IV)**

During the course of our work with pig APOBEC3 we were reminded that although pigs do not harbor a lentivirus, they have endogenous retroviruses that have received quite a bit of attention as a safety concern in xenotransplantation, where pigs are considered the most suitable donor to fill the need for transplantation material for humans. One of the major concepts coming out of APOBEC3 studies is that of species-specificity, APOBEC3 proteins seem to have a broad effect on viruses but successful viruses seem to have developed counter-measures against the APOBEC3 proteins of their host. This indicates that an APOBEC3 protein from a species to which the virus has not yet adapted may provide an effective strategy for thwarting this species-specific viral counter-defense. It is therefore reasonable to hypothesize that cross-species expression of an APOBEC3 protein may be used to create a powerful barrier to impede or perhaps even block retrovirus expression. Here, this rationale is applied to the specific question of whether human APOBEC3G expression can inhibit the transmission of PERV from pig to human cells.

### **A Co-culture Assay to Monitor PERV Transmission**

To determine whether expression of human APOBEC3G would inhibit the transfer of PERV from pig to human cells it was first necessary to establish a long-term co-culture system. A trans-well assay was set up to monitor PERV transmission from pig kidney PK-15 fibroblasts to recipient human embryonic kidney 293T cells (Figure 18A). These two cell types were used because transmission from PK-15 to 293 cells had been reported previously. The trans-well system enabled co-cultures to be sustained for several weeks, and it facilitated the recovery of each cell type for downstream analyses. An additional benefit of this co-culture system (not provided by transient assays) is that it enables the simultaneous analysis of multiple, endogenous PERV elements, which are precisely the targets one would want to monitor and ideally inhibit in (xeno)transplantation procedures. At each co-culture passage point, surplus human 293T cells were used to prepare genomic DNA. PERV transmission was monitored by subjecting these samples to quantitative (Q)-PCR. PERV-specific *pol* gene PCR products could be detected in the human genomic DNA samples after approximately two weeks of continuous co-culture (Figure 18B).

From the point of first detection onward, the total number of PERV transmissions continued to increase, averaging 190 new events per day per 50,000 cells ( $10^5$   $\beta$ -actin copies; SEM = 62;  $n = 5$  experiments). Importantly, pig cells did not breach the 293T cell compartment because Q-PCR analyses of the same human genomic DNA samples failed to detect a concomitant transfer of pig genomic DNA (Figure 18B). Moreover, PERV copy number did not increase over a two week interval when infected 293T cells were grown in isolation, indicating that PERV was not replicating in the human cells and that the majority of the observed transmission events were derived from the PK-15 cells (*i.e.*, new events). These results combine to indicate that the trans-well assay provides a robust system for monitoring *bona fide* zoonotic PERV transmissions.



**Figure 18.** PERV Transmission Assay. (A) Schematic of the co-culture system. PERV transmitting PK-15 cells are grown on top of the membrane of the insert and human 293T cells on the bottom of the well of the culture dish. Virus particles are depicted diffusing through membrane pores. (B) The zoonotic transmission of PERV from PK-15 cells to 293T cells is shown by the time-dependent accumulation of PERV pol gene DNA in the human cells (solid diamonds and squares). No concomitant transfer of pig genomic DNA occurred through the duration of these experiments (open diamonds and squares). This graph summarizes data for two independently derived PK-15 clones, V1 (squares) and V2 (diamonds). All data points were calculated using results from duplicate Q-PCR reactions of genomic DNA from three parallel (but independent) co-cultures. The error bars indicate the SEM.

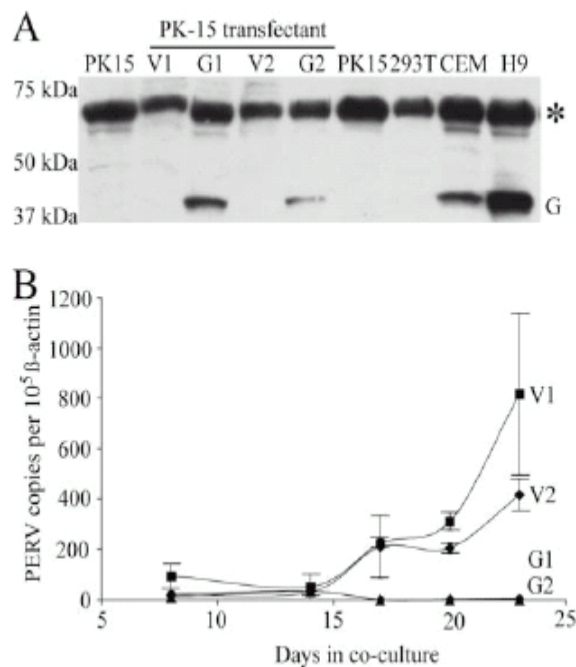
### Human APOBEC3G Inhibits PERV Transmission

The second key step in addressing our experimental question was isolating PK-15 clones that stably expressed human APOBEC3G. Clones expressing human APOBEC3G cDNA or an empty vector control were established in parallel (Figure 19A). Immunoblotting identified clones with APOBEC3G levels similar to those in

known APOBEC3G-expressing human T cell lines CEM and H9, which are non-permissive for growth of Vif-deficient HIV-1. Although it was impossible to achieve a physiologic expression level, the comparative immunoblot at least ensured that the levels of APOBEC3G were equal or lower than those present in well-studied, nonpermissive human T cell lines.

Co-culture experiments were set up to compare PERV transmission from two independently derived PK-15 clones expressing human APOBEC3G and two vector expressing controls. Remarkably, the human APOBEC3G expressing PK-15 clones showed levels of PERV transmission that were lower than the Q-PCR detection threshold of approximately 10 copies (Figure 19B). In contrast, the control clones showed high levels of PERV transfer by co-culture day 17 and transmission events continued to accumulate through the duration of the experiment.

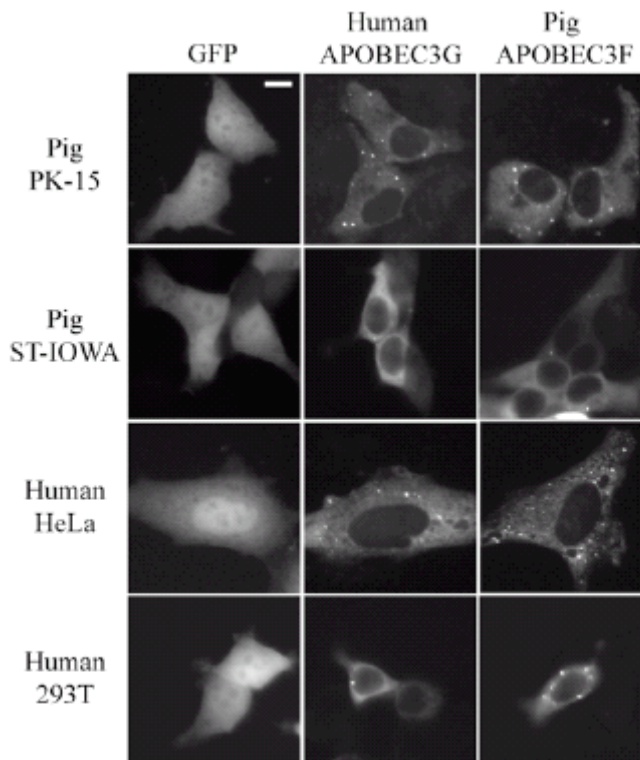
The kinetics of PERV transmission were similar to those reported in Figure 17B (these results also contributed to the aforementioned transfer rate calculations). These data were further corroborated by additional experiments where PERV transmission was monitored simultaneously by Q-PCR and by reverse-transcriptase ELISA assays.



**Figure 19.** Human APOBEC3G Inhibits PERV Transmission. (A) An immunoblot showing PK-15 clones expressing human APOBEC3G (G1 and G2) or a control vector (V1 and V2). PK-15 and 293T cell lysates were used as negative controls. CEM and H9 were used as positive controls for APOBEC3G expression. A non-specific (but pan-species) band is shown as a protein loading control (marked by an asterisk). (B) Q-PCR data using genomic DNA prepared from 293T cells co-cultured with two independently derived APOBEC3G expressing PK-15 clones (G1 and G2, circles and triangles, respectively) or two vector control clones (V1 and V2, diamonds and squares, respectively). The experimental parameters are identical to those used in Figure 1B.

### The Sub-cellular Distribution of Human APOBEC3G is Virtually Identical in Human and Pig Cell Lines

An ultimate application of the technology described here raises the potential problem that human APOBEC3G may not be subjected to (proper) post-translational regulation in pig cells and it may therefore promote carcinogenesis. Expression of human APOBEC3G in a heterologous system has been shown to trigger elevated levels of genomic C/G-to-T/A transition mutations. Therefore, to help mitigate this risk (in addition to establishing clones that expressed relatively modest APOBEC3G levels; above), we asked whether the predominantly cytoplasmic localization pattern of human APOBEC3G would be maintained in PK-15 and in a swine testes cell line, ST-IOWA (Figure 20). Unlike some other APOBEC3 proteins such as human APOBEC3B, which is mostly nuclear, both human APOBEC3G and pig APOBEC3F appeared predominantly cytoplasmic in either the human or the pig cell lines (Figure 20). Both proteins also appeared to concentrate in cytoplasmic punctae, which varied in number and were apparent in some of the cells regardless of species of origin. Overall, these near-identical localization patterns suggested that human APOBEC3G is not aberrantly regulated in pig cells and, interestingly, that these proteins might be subjected to the same cellular regulatory mechanism(s).



**Figure 20.** The Sub-cellular Distribution of Human APOBEC3G in Human and Pig Cell Lines. Sub-cellular distributions of GFP, human APOBEC3G-GFP and pig APOBEC3F-GFP in the indicated live pig and human cell lines. The scale bar in the top left panel indicates 10 mM. (figure courtesy of Mark D. Stenglein).

### **PERV Appears Resistant to Porcine APOBEC3Z2-Z3**

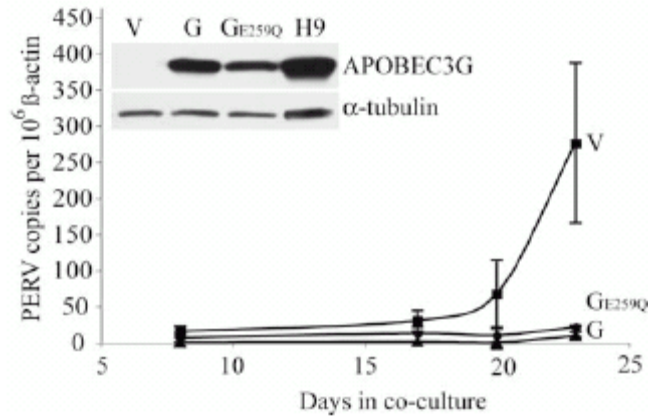
As demonstrated in paper II, pig APOBEC3Z2-Z3 can strongly inhibit the replication of HIV (regardless of Vif) and modestly inhibit the replication of MLV, a gamma-retrovirus phylogenetically related to PERV. Therefore, we wondered whether pig APOBEC3Z2-Z3 was expressed in PK-15 and, if so, whether PERV resists this cellular defense. To begin to address this possibility, RT-PCR was used to test PK-15 cells for pig APOBEC3Z2-Z3 expression. Pig *APOBEC3Z2-Z3* mRNA was detected readily. Full cDNA sequencing revealed that the predicted APOBEC3F protein of PK-15 cells was 98% similar to the variant we reported in paper II. Eight amino acid differences were found, but both the PK-15 and the previously reported APOBEC3Z2-Z3 sequences were represented in pig genomic DNA sequences suggesting that these may be breed-specific polymorphisms (paper III). These observations indicated that either PERV resists the endogenous APOBEC3Z2-Z3 protein of its host or that the level of suppression by pig APOBEC3Z2-Z3 is not sufficient to inhibit PERV transmission.

To begin to distinguish between these two hypotheses, PK-15 clones overexpressing pig APOBEC3Z2-Z3 were established and used in transmission experiments. The former hypothesis was favored because pig APOBEC3Z2-Z3 over-expression did not significantly interfere with PERV transmission. These results were further supported by PCR experiments showing that PERV could be amplified readily from 293T cells that had been co-cultured with PK-15 over-expressing pig APOBEC3Z2-Z3. These observations indicate that PERV is resistant to the endogenous APOBEC3 protein of its host.

### **Evidence that Human APOBEC3G Inhibits PERV by a Deamination-Independent Mechanism**

The hallmark of APOBEC3G-dependent retrovirus restriction is plus-strand G to A hypermutation, which is caused by the deamination of minus-strand cDNA C-to-U during reverse transcription. The deamination of cytosines within single-strand DNA requires glutamate 259 (E259) of APOBEC3G. To determine whether DNA deamination is required in the APOBEC3G-dependent inhibition of PERV transmission, we established a new set of PK-15 clones expressing APOBEC3G, APOBEC3G E259Q or a vector control (Figure 21, inset). Surprisingly, both APOBEC3G and the E259Q derivative diminished PERV transmission to near

background levels (Figure 21). These data demonstrated that the mechanism of inhibition does not require the DNA deaminase activity of APOBEC3G. These data were further supported by the fact that plus strand G-toA hypermutations were not apparent in the DNA of the rare transmission events that occurred in the presence of human APOBEC3G.



**Figure 21.** The APOBEC3G-dependent Inhibition of PERV Transmission is Deamination-Independent. PERV-specific Q-PCR data using genomic DNA prepared from 293T cells co-cultured with PK-15 clones expressing APOBEC3G (G; triangles), APOBEC3G-E259Q (GE259Q; circles) or empty vector (V; squares). Two datasets, each with an independent PK-15 clone in three replica co-culture wells, were collected in parallel and averaged for each data point. One standard error of the mean is shown. The experimental parameters are identical to those used in Figure 1B. The inset immunoblots show the APOBEC3G and  $\alpha$ -tubulin levels of representative PK-15 clones expressing the indicated construct. The human T cell line H9 provided a positive control for APOBEC3G expression.

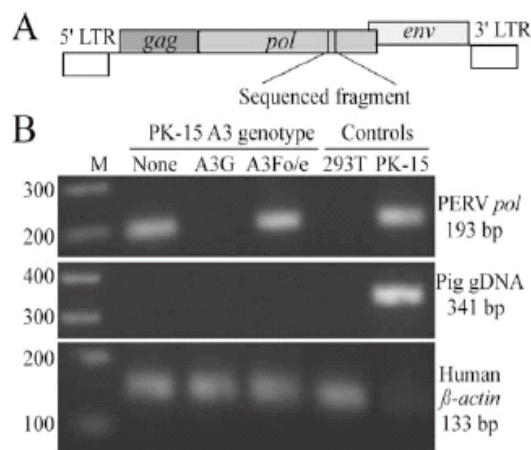
### Genetic Variation in Zoonosed PERV DNA Sequences

To begin to genotype the infectious PERVs and to further probe the mechanism of PERV restriction by human APOBEC3G, the PERV *pol* gene DNA was amplified from human 293T cells, cloned and sequenced (Figure 22A). Twenty-nine and twenty-two sequences were analysed from APOBEC3G and control experiments, respectively. To minimize possible PCR biases, any sequence that was recovered multiple times was considered one event, unless it arose from independent experiments. These DNA sequence analyses revealed several important points.

First, in contrast to vector control and pig APOBEC3F over-expressing cocultures, the PERV *pol* gene DNA was difficult to amplify from the genomic DNA of 293T cells that had been co-cultured with APOBEC3G-expressing PK-15 cells (Figure 22B). Taking this together with the observation that APOBEC3G does not affect PK-

15 virus production (similar RT levels were observed in cell-free supernatants in the presence or absence of APOBEC3G), we infer that APOBEC3G restricts PERV transmission after virus production but before provirus integration (*i.e.*, between entry and integration). APOBEC3G may restrict PERV at an early reverse transcription stage, possibly by interfering with primer binding, DNA synthesis and/or integration as shown recently for APOBEC3G and HIV-1 substrates.

Second, control co-culture PERV transmission events were exceptionally diverse, as 15 unique *pol* sequences were detected and only 3 were found multiple times (Figure 22B). These data suggested that PK-15 cells have at least 15 active PERVs capable of infecting human 293T cells, a number consistent with previous studies that reported the existence of approximately 17-50 PERV copies in total (with only a fraction being replication-competent). In contrast, the rare PERV sequences derived from the APOBEC3G co-culture experiments showed a much lower genetic complexity. Only three unique sequences were recovered, each differing by a single nucleotide. In parallel experiments with HIV-based viruses, this APOBEC3G expression construct caused approximately 30 G-to-A hypermutations per 1000 bases analyzed. Thus, approximately 12 G-to-A transitions should have been recovered in these PERV DNA analyses (nearly 90 if multiply recovered sequences would have been considered). The absence of hypermutated PERV proviral DNA provided further support for a deaminase-independent mechanism of restriction, which may share features with other instances described previously.



**Figure 22.** Genetic Variation in Zoonosed PERV DNA Sequences. (A) A schematic of the PERV genome indicating the relevant 193 bp *pol* gene fragment (B) An agarose gel showing that PERV DNA was amplified readily from 293T cells cultured with PK-15 control clones (None) and pig APOBEC3F overexpressing clones (A3Fo/e) but not with PK-15 clones expressing human APOBEC3G (A3G; top panel). PERV *pol* gene DNA (top panel) and pig genomic DNA (APOBEC3F locus; middle panel).

## 5. DISCUSSION

In recent years it has become increasingly clear that mammalian cells possess a variety of antiretroviral activities, not only to combat the attack of exogenous retroviruses, but also to defend themselves against possible harm caused by transposition of endogenous retroelements which constitute a large portion of the genome of all mammalian species. The APOBEC3 protein family of cytosine deaminases has played a central role in this recently emerging experimental field of “retrovirus restriction factors”. In the work presented in this thesis the main focus has been on the cloning and characterization of the APOBEC3 proteins of sheep and two other closely related artiodactyl species, cattle and pigs.

The first part of the thesis described in paper I focused on delineating the genetic components of the difference in phenotype between the two MVV molecular clones KV1772 and KS1. The attenuated replication and lack of infectivity of the KS1 clone was traced to simultaneous mutations, one in the CA domain of gag and the other in Vif. The CA mutation affected replication in macrophages somewhat, but with the addition of the *vif* mutation the replication in macrophages was markedly reduced. The mutation in *vif* on its own did not affect the replication in macrophages. During the course of this work it became apparent that the main role of the Vif protein of HIV-1 was to battle the human protein APOBEC3G, a member of a newly described family of cytosine deaminases (Harris et al., 2003a; Mangeat et al., 2003; Sheehy et al., 2002; Zhang et al., 2003). Although the interplay between the CA and Vif mutations in MVV may not be related to the role of Vif against these deaminases we decided to venture into the new and rapidly expanding field of the antiretroviral effects of the APOBEC3 family of cytosine deaminases.

The initial aim of these studies was to clone and characterize the APOBEC3 proteins of sheep. Due to the fact that little sheep DNA sequence exists in public databases, the project was broadened to include other closely related species of the *artiodactyla* order of placental mammals. These studies focused on APOBEC3 proteins from three representative artiodactyl species: Cow (*Bos taurus*), sheep (*Ovis aries*) and pigs (*Sus scrofa*), that were predicted to be of similar size and domain organization as the well characterized human APOBEC3F and APOBEC3G, that are

the most potent anti-retroviral human APOBEC3 proteins, as well as the single mouse APOBEC3 protein.

The cloned double domain APOBEC3Z2-Z3 proteins from these three cloven hoof ungulates were all active deaminases and robustly inhibited the activity of HIV-1 regardless of the expression of Vif. They also showed moderate inhibitory activity against the simple gammaretrovirus MLV. One of the more intriguing features of the artiodactyl A3Z2-Z3 proteins is the fact that the DNA cytosine deaminase activity resides in the N-terminal zinc-binding domain, and not in the C-terminal zinc-binding domain like several of the human double domain A3 proteins. These data highlight the modular nature of the A3 proteins, which can clearly function as single domain proteins, as double domain proteins with a catalytically active C-terminal domain, as double domain proteins with a catalytically active N-terminal zinc-binding domain or as two distinct single domain proteins (paper III). Prior studies had demonstrated that the N-terminal, zinc-binding domain of HsA3G mediates interactions with both HIV-1 Vif and Gag (Alce and Popik, 2004; Cen et al., 2004; Conticello et al., 2003; Luo et al., 2004; Schafer et al., 2004; Svarovskaia et al., 2004; Zennou et al., 2004). The C-terminal domain of HsA3G dictates DNA cytosine deaminase activity, and both domains contribute to dimerization activity (Hache et al., 2005; Newman et al., 2005; Shindo et al., 2003). Thus, the division of activities between the N- and the C-terminal domains suggested that the A3 proteins are rapidly evolving as modular domains centered upon the conserved, zinc-binding deaminase motif.

The fact that mammals have varying numbers of APOBEC3 genes, with humans having seven and mice only having one, led us to wonder whether there were additional APOBEC3 genes in the artiodactyls to those we had worked with in paper II. To address this point and to learn more about the evolution and functionality of APOBEC3 genes in mammals, we sequenced and characterized the full APOBEC3 repertoire of sheep and pigs and compared that to available information from the cow (paper III). We demonstrated that sheep and cattle have three *A3* genes, *A3Z1*, *A3Z2* and *A3Z3*. However, the latter two genes and their counterpart in pigs have the unique ability to produce a double-domain protein A3Z2-Z3 (which was the proteins initially cloned and characterized in paper II), in addition to single-domain polypeptides. Thus, the A3 proteome of these species is more formidable than gene number alone would indicate. These studies also helped highlight the important point

that, although A3 proteins consist of either one or two conserved Z domains, each of these domains can function and evolve independently.

Prior to the present studies, it was clear that most (if not all) placental mammals had Z2- and Z3-type A3 domains (*e.g.*, human, mouse, cat, pig, sheep and cow). It was far less clear how broadly the Z1 domain distributed. In paper III two critical lines of evidence were presented, strongly indicating that the Z1 distribution is equally broad and, importantly, that the common ancestor of placental mammals had a Z1-Z2-Z3 A3 gene repertoire, similar to that of present-day sheep and cattle. These data support a model in which the common ancestor of the primate- and the artiodactyl-containing mammalian super-orders, Euarchontoglires and Laurasiatheria, respectively, had a A3Z1 gene and precisely one of each of the three conserved Z domain types (*i.e.*, a divergent model for A3 gene evolution, as opposed to one in which A3Z1 genes evolved independently in several limbs of the mammalian tree). It is noteworthy that our pan-species Z1 PCR analyses failed to generate product from opossum genomic DNA and that the recently released opossum and platypus genomic sequences lack A3 genes. This is unlikely to be a gap in the DNA sequence assemblies because, like non-mammalian vertebrates, DNA and protein searches clearly revealed the A3-flanking genes *CBX6* and *CBX7* in both animals. Thus, unfortunately, these two interesting non-placental mammals are unlikely to provide significant insights into the earliest stages of A3 gene evolution (*i.e.*, pre dating the Z1-Z2-Z3 ancestor described here). Perhaps data from the other two placental mammal super-orders, Afrotheria and Xenarthra (*e.g.*, represented by animals such as elephants and armadillos, respectively), will help shed light on earlier stages of A3 gene evolution, when presumably an *AID*-like gene transposed between *CBX6* and *CBX7* and duplicated to give rise to the ancestral Z1-Z2-Z3 locus. The elephants and armadillos are specifically used as examples of the two super-orders since ongoing genome projects exist for both of these animal species and it will be very interesting to see what the APOBEC3 loci of these species look like or even if they contain any A3 genes. This shows as well that APOBEC1 is older than the APOBEC3, since it has been previously shown that opossums encode a functional APOBEC1 protein. Nevertheless, because all current data indicate that the A3 genes are specific to placental mammals; it can be hypothesized that a unique role of these genes may relate to the placenta itself, where the A3 proteins may function

to help protect the developing fetus from potentially harmful retrotransposition events and/or retroviral infections.

While working with the APOBEC3 proteins of pigs, we were reminded that they were a species that do not harbor lentiviruses, however they have been under scrutiny based on the fact that they have endogenous retroviruses (PERV) that have been shown to be able to be transmitted from pig to human cells under laboratory conditions. That caused some concern for the use of pig materials as resources for transplantation purposes, where pigs have been proposed as the most convenient animal, due to their human-like physiology, large litter sizes and how easy they are to breed. We decided to try a novel approach to inhibit the transmission of PERVs from pig to human cells, by introducing a “foreign” APOBEC3 protein into the PERV producing pig cells to see whether its expression would impede the transfer of the viruses from the pig cells to the human cells. To do this we established a quantitative assay to monitor the zoonotic transmission of PERV to human 293T cells, using a co-culture set-up and quantitative real-time PCR. Expression of human APOBEC3G in the pig PK-15 cell line strongly inhibited PERV zoonoses, while the endogenous APOBEC3Z2-Z3 protein of pigs appeared considerably less effective. These data were the first to demonstrate that APOBEC3 proteins can be used purposefully to reduce if not prevent zoonotic retroviral infections.

Our data indicate that the engineering of pigs to express human APOBEC3G (or an equally potent non-porcine APOBEC3) may result in animals whose cells and tissues are much less likely to disseminate functional PERV. The deamination-independence of the restriction mechanism suggests that a catalytically inert APOBEC3G protein, such as E259Q, may be equally potent and simultaneously reduce the risk of cancer-promoting mutagenesis. APOBEC3G or APOBEC3G-E295Q expressing pigs may therefore constitute safer source animals for pig-to-human xenotransplantation procedures. In contrast to knockdown, knockout (by gene targeting or selective breeding) or most chemical-based anti-viral approaches to neutralize PERV, the APOBEC3 anti-viral defense system has several advantages including a potentially broad neutralizing activity and an applicability to situations where many copies of a virus are already present in a genome. Analogous transgenic applications can be envisaged, such as using cross-species APOBEC3 expression to purposefully impede known viruses (*e.g.*, the AIDS virus HIV-1 or the Hepatitis B virus HBV). Moreover, for humans and other mammals with multiple APOBEC3

proteins, our data encourage the development of methods to induce/up-regulate endogenous APOBEC3 proteins, which have the capacity but may not normally restrict a particular virus (*e.g.*, human APOBEC3B and HIV-1). Our findings were strengthened as an alternative in the prevention of PERV transmission when in a recent paper PERV A and A/C viruses were shown to be insensitive to the inhibitory effect of a range of Trim5 $\alpha$  proteins including human Trim5 $\alpha$  (Wood et al., 2009).

Based on our work with PERV it is likely that APOBEC3 proteins could be used therapeutically. Anti-retroviral therapies might take advantage of this powerful innate defense system. In the fight against HIV-1 and AIDS in the near future the focus will most likely be on the use of small molecules to disrupt the APOBEC3:Vif interaction (Nathans et al., 2008).

Many key questions remain unanswered both on the mechanism and the evolution of the APOBEC3 proteins. It will be interesting to gain knowledge of the APOBEC3 repertoire of more mammalian species especially from other sub-orders yet to be sequenced as well as mammalian species with special adaptations to different or extreme environments such as whales and other marine mammals. There are also questions to be answered on the interaction between the Vif proteins of non-primate lentiviruses and the APOBEC3 proteins of their hosts. Studies have shown that MVV lacking the *vif* gene is subject to G to A hypermutations (Kristbjornsdottir et al., 2004) yet the Vif proteins of non-primate lentiviruses lack the H<sub>x5</sub>C<sub>x17-18</sub>C<sub>x3-5</sub>H zinc-binding domain that mediates Cul5 binding. The Vif proteins of the non-primate lentiviruses have therefore adopted a different approach than the primate lentiviral Vif proteins in quenching the threat of the host APOBEC3 proteins. An especially interesting phenomenon is how the equine lentivirus EIAV avoids the multiple APOBEC3 proteins found in horses since it does not harbor a Vif protein.

Further studies could bring further understanding in virus host interactions and the evolution and adaptation of host restriction factors as well as the role of species specificity in retroviral infections.

## 6. REFERENCES

- Abudu, A., Takaori-Kondo, A., Izumi, T., Shirakawa, K., Kobayashi, M., Sasada, A., Fukunaga, K. and Uchiyama, T. (2006) Murine retrovirus escapes from murine APOBEC3 via two distinct novel mechanisms. *Curr Biol* 16(15), 1565-70.
- Acheampong, E., Rosario-Otero, M., Dornburg, R. and Pomerantz, R.J. (2003) Replication of lentiviruses. *Front Biosci* 8, s156-74.
- Agnarsdottir, G., Thorsteinsdottir, H., Oskarsson, T., Matthiasdottir, S., St Haflidadottir, B., Andresson, O.S. and Andresdottir, V. (2000) The long terminal repeat is a determinant of cell tropism of maedi- visna virus. *J Gen Virol* 81, 1901-5.
- Akiyoshi, D.E., Denaro, M., Zhu, H., Greenstein, J.L., Banerjee, P. and Fishman, J.A. (1998) Identification of a full-length cDNA for an endogenous retrovirus of miniature swine. *J Virol* 72(5), 4503-7.
- Alce, T.M. and Popik, W. (2004) APOBEC3G is incorporated into virus-like particles by a direct interaction with HIV-1 Gag nucleocapsid protein. *J Biol Chem* 279(33), 34083-6.
- Alkhatib, G., Combadiere, C., Broder, C.C., Feng, Y., Kennedy, P.E., Murphy, P.M. and Berger, E.A. (1996) CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* 272(5270), 1955-8.
- Anant, S. and Davidson, N.O. (2001) Molecular mechanisms of apolipoprotein B mRNA editing. *Curr Opin Lipidol* 12(2), 159-65.
- Andresson, O.S., Elser, J.E., Tobin, G.J., Greenwood, J.D., Gonda, M.A., Georgsson, G., Andresdottir, V., Benediktsdottir, E., Carlsdottir, H.M. and Mantyla, E.O. (1993) Nucleotide sequence and biological properties of a pathogenic proviral molecular clone of neurovirulent visna virus. *Virology* 193(1), 89-105.
- Armstrong, J.A., Porterfield, J.S. and De Madrid, A.T. (1971) C-type virus particles in pig kidney cell lines. *J Gen Virol* 10(2), 195-8.
- Bailey, J.A., Carrel, L., Chakravarti, A. and Eichler, E.E. (2000) Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc Natl Acad Sci U S A* 97(12), 6634-9.
- Baltimore, D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 226(5252), 1209-11.
- Baust, C., Baillie, G.J. and Mager, D.L. (2002) Insertional polymorphisms of ETn retrotransposons include a disruption of the wiz gene in C57BL/6 mice. *Mamm Genome* 13(8), 423-8.
- Beale, R.C., Petersen-Mahrt, S.K., Watt, I.N., Harris, R.S., Rada, C. and Neuberger, M.S. (2004) Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J Mol Biol* 337(3), 585-96.
- Bininda-Emonds, O.R., Cardillo, M., Jones, K.E., MacPhee, R.D., Beck, R.M., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L. and Purvis, A. (2007) The delayed rise of present-day mammals. *Nature* 446(7135), 507-12.
- Bishop, K.N., Holmes, R.K. and Malim, M.H. (2006) Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. *J Virol* 80(17), 8450-8.
- Bishop, K.N., Holmes, R.K., Sheehy, A.M. and Malim, M.H. (2004) APOBEC-mediated editing of viral RNA. *Science* 305(5684), 645.

- Bishop, K.N., Verma, M., Kim, E.Y., Wolinsky, S.M. and Malim, M.H. (2008) APOBEC3G inhibits elongation of HIV-1 reverse transcripts. *PLoS Pathog* 4(12), e1000231.
- Bjarnadottir, H., Gudmundsson, B., Gudnason, J. and Jonsson, J.J. (2006) Encapsidation determinants located downstream of the major splice donor in the maedi-visna virus leader region. *J Virol* 80(23), 11743-55.
- Blusch, J.H., Patience, C. and Martin, U. (2002) Pig endogenous retroviruses and xenotransplantation. *Xenotransplantation* 9(4), 242-51.
- Boeke, J.D. and Stoye, J.P. (1997) Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In: J.M. Coffin, S.H. Hughes and H.E. Varmus (Eds), *Retroviruses*, pp. 343-436. Cold Spring Harbor Laboratory Press.
- Bogerd, H.P. and Cullen, B.R. (2008) Single-stranded RNA facilitates nucleocapsid: APOBEC3G complex formation. *Rna* 14(6), 1228-36.
- Bogerd, H.P., Tallmadge, R.L., Oaks, J.L., Carpenter, S. and Cullen, B.R. (2008) Equine infectious anemia virus resists the antiretroviral activity of equine APOBEC3 proteins through a packaging-independent mechanism. *J Virol* 82(23), 11889-901.
- Bosch, S., Arnault, C. and Jestin, A. (2000) Study of full-length porcine endogenous retrovirus genomes with envelope gene polymorphism in a specific-pathogen-free Large White swine herd. *J Virol* 74(18), 8575-81.
- Bouyac, M., Courcoul, M., Bertoia, G., Baudat, Y., Gabuzda, D., Blanc, D., Chazal, N., Boulanger, P., Sire, J., Vigne, R. and Spire, B. (1997a) Human immunodeficiency virus type 1 Vif protein binds to the Pr55Gag precursor. *J Virol* 71(12), 9358-65.
- Bouyac, M., Rey, F., Nascimbeni, M., Courcoul, M., Sire, J., Blanc, D., Clavel, F., Vigne, R. and Spire, B. (1997b) Phenotypically Vif- human immunodeficiency virus type 1 is produced by chronically infected restrictive cells. *J Virol* 71(3), 2473-7.
- Breese, S.S., Jr. (1970) Virus-like particles occurring in cultures of stable pig kidney cell lines. Brief report. *Arch Gesamte Virusforsch* 30(4), 401-4.
- Brewer, L.A., Lwamba, H.C., Murtaugh, M.P., Palmenberg, A.C., Brown, C. and Njenga, M.K. (2001) Porcine encephalomyocarditis virus persists in pig myocardium and infects human myocardial cells. *J Virol* 75(23), 11621-9.
- Brodie, R., Smith, A.J., Roper, R.L., Tcherepanov, V. and Upton, C. (2004) Base-By-Base: single nucleotide-level analysis of whole viral genome alignments. *BMC Bioinformatics* 5, 96.
- Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V. and Kazazian, H.H., Jr. (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* 100(9), 5280-5.
- Brown, P.O. (1997) Integration. In: J.M. Coffin, S.H. Hughes and H.E. Varmus (Eds), *Retroviruses*, pp. 161-204. Cold Spring Harbor Laboratory Press.
- Carruth, L.M., Morse, B.A. and Clements, J.E. (1996) The leucine domain of the visna virus Tat protein mediates targeting to an AP-1 site in the viral long terminal repeat. *J Virol* 70(7), 4338-44.
- Cen, S., Guo, F., Niu, M., Saadatmand, J., Deflassieux, J. and Kleiman, L. (2004) The interaction between HIV-1 Gag and APOBEC3G. *J Biol Chem* 279(32), 33177-84.

- Chatterji, U., Grant, C.K. and Elder, J.H. (2000) Feline immunodeficiency virus Vif localizes to the nucleus. *J Virol* 74(6), 2533-40.
- Chelico, L., Pham, P., Calabrese, P. and Goodman, M.F. (2006) APOBEC3G DNA deaminase acts processively 3' → 5' on single-stranded DNA. *Nat Struct Mol Biol* 13(5), 392-9.
- Chen, K.M., Harjes, E., Gross, P.J., Fahmy, A., Lu, Y., Shindo, K., Harris, R.S. and Matsuo, H. (2008) Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* 452(7183), 116-9.
- Cherepanov, P. (2007) LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic Acids Res* 35(1), 113-24.
- Chiu, Y.L. and Greene, W.C. (2008) The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu Rev Immunol* 26, 317-53.
- Chiu, Y.L., Soros, V.B., Kreisberg, J.F., Stopak, K., Yonemoto, W. and Greene, W.C. (2005) Cellular APOBEC3G restricts HIV-1 infection in resting CD4<sup>+</sup> T cells. *Nature* 435(7038), 108-14.
- Chowdhury, I.H., Chao, W., Potash, M.J., Sova, P., Gendelman, H.E. and Volsky, D.J. (1996) vif-negative human immunodeficiency virus type 1 persistently replicates in primary macrophages, producing attenuated progeny virus. *J Virol* 70(8), 5336-45.
- Chua, K.B. (2003) Nipah virus outbreak in Malaysia. *J Clin Virol* 26(3), 265-75.
- Chua, K.B., Goh, K.J., Wong, K.T., Kamarulzaman, A., Tan, P.S., Ksiazek, T.G., Zaki, S.R., Paul, G., Lam, S.K. and Tan, C.T. (1999) Fatal encephalitis due to Nipah virus among pig-farmers in Malaysia. *Lancet* 354(9186), 1257-9.
- Clark, D.A., Fryer, J.F., Tucker, A.W., McArdle, P.D., Hughes, A.E., Emery, V.C. and Griffiths, P.D. (2003) Porcine cytomegalovirus in pigs being bred for xenograft organs: progress towards control. *Xenotransplantation* 10(2), 142-8.
- Clemenceau, B., Lalain, S., Martignat, L. and Sai, P. (1999) Porcine endogenous retroviral mRNAs in pancreas and a panel of tissues from specific pathogen-free pigs. *Diabetes Metab* 25(6), 518-25.
- Clements, J.E. and Zink, M.C. (1996) Molecular biology and pathogenesis of animal lentivirus infections. *Clin Microbiol Rev* 9(1), 100-17.
- Coffin, J.M. (1992) Structure and classification of retroviruses. In: J.A. Levy (Ed), *The Retroviridae*, pp. 19-44. Vol. 1. Plenum Press, New York.
- Conticello, S.G., Harris, R.S. and Neuberger, M.S. (2003) The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr Biol* 13(22), 2009-13.
- Conticello, S.G., Thomas, C.J., Petersen-Mahrt, S.K. and Neuberger, M.S. (2005) Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* 22(2), 367-77.
- Cooper, D.K. (2003) Clinical xenotransplantation--how close are we? *Lancet* 362(9383), 557-9.
- Cordier, G., Cozon, G., Greenland, T., Rocher, F., Guiguen, F., Guerret, S., Brune, J. and Mornex, J.F. (1990) In vivo activation of alveolar macrophages in ovine lentivirus infection. *Clin Immunol Immunopathol* 55(3), 355-67.
- Courcoul, M., Patience, C., Rey, F., Blanc, D., Harmache, A., Sire, J., Vigne, R. and Spire, B. (1995) Peripheral blood mononuclear cells produce normal amounts of defective Vif- human immunodeficiency virus type 1 particles which are restricted for the preretrotranscription steps. *J Virol* 69(4), 2068-74.

- Cramer, D.V. (2000) Natural antibodies and the host immune responses to xenografts. *Xenotransplantation* 7(2), 83-92.
- Cullen, B.R. (2006) Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors. *J Virol* 80(3), 1067-76.
- Dalgleish, A.G. (1985) HTLV-I/II and T-cell proliferation. *Lancet* 2(8467), 1302.
- Dalgleish, A.G., Beverley, P.C., Clapham, P.R., Crawford, D.H., Greaves, M.F. and Weiss, R.A. (1984) The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* 312(5996), 763-7.
- Dang, Y., Siew, L.M. and Zheng, Y.H. (2008) APOBEC3G is degraded by the proteasomal pathway in a Vif-dependent manner without being polyubiquitylated. *J Biol Chem* 283(19), 13124-31.
- Dang, Y., Wang, X., Esselman, W.J. and Zheng, Y.H. (2006) Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. *J Virol*.
- Deininger, P.L. and Batzer, M.A. (1999) Alu repeats and human disease. *Mol Genet Metab* 67(3), 183-93.
- Deininger, P.L. and Batzer, M.A. (2002) Mammalian retroelements. *Genome Res* 12(10), 1455-65.
- Delebecque, F., Suspene, R., Calattini, S., Casartelli, N., Saib, A., Froment, A., Wain-Hobson, S., Gessain, A., Vartanian, J.P. and Schwartz, O. (2006) Restriction of foamy viruses by APOBEC cytidine deaminases. *J Virol* 80(2), 605-14.
- Derse, D., Hill, S.A., Princler, G., Lloyd, P. and Heidecker, G. (2007) Resistance of human T cell leukemia virus type 1 to APOBEC3G restriction is mediated by elements in nucleocapsid. *Proc Natl Acad Sci U S A* 104(8), 2915-20.
- Desrosiers, R.C., Lifson, J.D., Gibbs, J.S., Czajak, S.C., Howe, A.Y., Arthur, L.O. and Johnson, R.P. (1998) Identification of highly attenuated mutants of simian immunodeficiency virus. *J Virol* 72(2), 1431-7.
- Doehle, B.P., Schafer, A. and Cullen, B.R. (2005a) Human APOBEC3B is a potent inhibitor of HIV-1 infectivity and is resistant to HIV-1 Vif. *Virology* 339(2), 281-8.
- Doehle, B.P., Schafer, A., Wiegand, H.L., Bogerd, H.P. and Cullen, B.R. (2005b) Differential sensitivity of murine leukemia virus to APOBEC3-mediated inhibition is governed by virion exclusion. *J Virol* 79(13), 8201-7.
- Dorak, M.T. (2006) Real-time PCR. Taylor and Francis Group, New York.
- Dutko, J.A., Schafer, A., Kenny, A.E., Cullen, B.R. and Curcio, M.J. (2005) Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases. *Curr Biol* 15(7), 661-6.
- Ehlers, B., Ulrich, S. and Goltz, M. (1999) Detection of two novel porcine herpesviruses with high similarity to gammaherpesviruses. *J Gen Virol* 80 (Pt 4), 971-8.
- Eto, T., Kinoshita, K., Yoshikawa, K., Muramatsu, M. and Honjo, T. (2003) RNA-editing cytidine deaminase Apobec-1 is unable to induce somatic hypermutation in mammalian cells. *Proc Natl Acad Sci U S A* 100(22), 12895-8.
- Fahrenkrug, S.C., Smith, T.P., Freking, B.A., Cho, J., White, J., Vallet, J., Wise, T., Rohrer, G., Pertea, G., Sultana, R., Quackenbush, J. and Keele, J.W. (2002) Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly. *Mamm Genome* 13(8), 475-8.
- Freed, E.O. (2001) HIV-1 replication. *Somat Cell Mol Genet* 26(1-6), 13-33.

- Fryer, J.F., Griffiths, P.D., Fishman, J.A., Emery, V.C. and Clark, D.A. (2001) Quantitation of porcine cytomegalovirus in pig tissues by PCR. *J Clin Microbiol* 39(3), 1155-6.
- Gabuzda, D.H., Lawrence, K., Langhoff, E., Terwilliger, E., Dorfman, T., Haseltine, W.A. and Sodroski, J. (1992) Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes. *J Virol* 66(11), 6489-95.
- Gabuzda, D.H., Li, H., Lawrence, K., Vasir, B.S., Crawford, K. and Langhoff, E. (1994) Essential role of vif in establishing productive HIV-1 infection in peripheral blood T lymphocytes and monocyte/macrophages. *J Acquir Immune Defic Syndr* 7(9), 908-15.
- Gibbs, J.S., Regier, D.A. and Desrosiers, R.C. (1994) Construction and in vitro properties of SIVmac mutants with deletions in "nonessential" genes. *AIDS Res Hum Retroviruses* 10(5), 607-16.
- Gifford, R., Kabat, P., Martin, J., Lynch, C. and Tristem, M. (2005) Evolution and distribution of class II-related endogenous retroviruses. *J Virol* 79(10), 6478-86.
- Goncalves, J., Jallepalli, P. and Gabuzda, D.H. (1994) Subcellular localization of the Vif protein of human immunodeficiency virus type 1. *J Virol* 68(2), 704-12.
- Goncalves, J., Shi, B., Yang, X. and Gabuzda, D. (1995) Biological activity of human immunodeficiency virus type 1 Vif requires membrane targeting by C-terminal basic domains. *J Virol* 69(11), 7196-204.
- Gonda, M.A., Wong-Staal, F., Gallo, R.C., Clements, J.E., Narayan, O. and Gilden, R.V. (1985) Sequence homology and morphologic similarity of HTLV-III and visna virus, a pathogenic lentivirus. *Science* 227(4683), 173-7.
- Goodier, J.L., Ostertag, E.M., Du, K. and Kazazian, H.H., Jr. (2001) A novel active L1 retrotransposon subfamily in the mouse. *Genome Res* 11(10), 1677-85.
- Gudmundsson, B., Jónsson, S.R., Ólafsson, O., Agnarsdóttir, G., Matthíasdóttir, S., Georgsson, G., Torsteinsdóttir, S., Svansson, V., Kristbjornsdóttir, H.B., Franzdóttir, S., Andrésson, Ó.S. and Andrésdóttir, V. (2005) Simultaneous mutations in CA and Vif of maedi-visna virus cause attenuated replication in macrophages and reduced infectivity in vivo. *J Virol* 79, 15038-15042.
- Gudnadottir, M. (1974) Visna-maedi in sheep. *Prog Med Virol* 18(0), 336-49.
- Guo, F., Cen, S., Niu, M., Saadatmand, J. and Kleiman, L. (2006) Inhibition of formula-primed reverse transcription by human APOBEC3G during human immunodeficiency virus type 1 replication. *J Virol* 80(23), 11710-22.
- Haase, A.T. (1975) The slow infection caused by visna virus. *Curr Top Microbiol Immunol* 72, 101-56.
- Hache, G., Liddament, M.T. and Harris, R.S. (2005) The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. *J Biol Chem* 280(12), 10920-4.
- Haflidadottir, B.S., Matthíasdottir, S., Agnarsdottir, G., Torsteinsdottir, S., Petursson, G., Andreesson, O.S. and Andresdottir, V. (2008) Mutational analysis of a principal neutralization domain of visna/maedi virus envelope glycoprotein. *J Gen Virol* 89(Pt 3), 716-21.
- Halbur, P.G., Kasorndorkbua, C., Gilbert, C., Guenette, D., Potters, M.B., Purcell, R.H., Emerson, S.U., Toth, T.E. and Meng, X.J. (2001) Comparative pathogenesis of infection of pigs with hepatitis E viruses recovered from a pig and a human. *J Clin Microbiol* 39(3), 918-23.

- Harjes, E., Gross, P.J., Chen, K.M., Lu, Y., Shindo, K., Nowarski, R., Gross, J.D., Kotler, M., Harris, R.S. and Matsuo, H. (2009) An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. *J Mol Biol* 389(5), 819-32.
- Harmache, A., Bouyac, M., Audoly, G., Hieblot, C., Peveri, P., Vigne, R. and Suzan, M. (1995) The vif gene is essential for efficient replication of caprine arthritis encephalitis virus in goat synovial membrane cells and affects the late steps of the virus replication cycle. *J Virol* 69(6), 3247-57.
- Harmache, A., Russo, P., Guiguen, F., Vitu, C., Vignoni, M., Bouyac, M., Hieblot, C., Pepin, M., Vigne, R. and Suzan, M. (1996) Requirement of caprine arthritis encephalitis virus vif gene for in vivo replication. *Virology* 224(1), 246-55.
- Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S. and Malim, M.H. (2003a) DNA deamination mediates innate immunity to retroviral infection. *Cell* 113(6), 803-9.
- Harris, R.S. and Liddament, M.T. (2004) Retroviral restriction by APOBEC proteins. *Nat Rev Immunol* 4(11), 868-77.
- Harris, R.S., Petersen-Mahrt, S.K. and Neuberger, M.S. (2002) RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Molecular Cell* 10(5), 1247-53.
- Harris, R.S., Sheehy, A.M., Craig, H.M., Malim, M.H. and Neuberger, M.S. (2003b) DNA deamination: not just a trigger for antibody diversification but also a mechanism for defense against retroviruses. *Nat Immunol* 4(7), 641-3.
- Hess, J.L., Clements, J.E. and Narayan, O. (1985) cis- and trans-acting transcriptional regulation of visna virus. *Science* 229(4712), 482-5.
- Hess, J.L., Small, J.A. and Clements, J.E. (1989) Sequences in the visna virus long terminal repeat that control transcriptional activity and respond to viral trans-activation: involvement of AP-1 sites in basal activity and trans-activation. *J Virol* 63(7), 3001-15.
- Holmes, R.K., Malim, M.H. and Bishop, K.N. (2007) APOBEC-mediated viral restriction: not simply editing? *Trends Biochem Sci* 32(3), 118-28.
- Huthoff, H. and Malim, M.H. (2007) Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation. *J Virol* 81(8), 3807-15.
- Inoshima, Y., Kohmoto, M., Ikeda, Y., Yamada, H., Kawaguchi, Y., Tomonaga, K., Miyazawa, T., Kai, C., Umemura, T. and Mikami, T. (1996) Roles of the auxiliary genes and AP-1 binding site in the long terminal repeat of feline immunodeficiency virus in the early stage of infection in cats. *J Virol* 70(12), 8518-26.
- Irgang, M., Sauer, I.M., Karlas, A., Zeilinger, K., Gerlach, J.C., Kurth, R., Neuhaus, P. and Denner, J. (2003) Porcine endogenous retroviruses: no infection in patients treated with a bioreactor based on porcine liver cells. *J Clin Virol* 28(2), 141-54.
- Iwatani, Y., Takeuchi, H., Strebel, K. and Levin, J.G. (2006) Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. *J Virol* 80(12), 5992-6002.
- Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J. and Navaratnam, N. (2002) An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* 79(3), 285-96.

- Jonsson, S.R., Hache, G., Stenglein, M.D., Fahrenkrug, S.C., Andresdottir, V. and Harris, R.S. (2006) Evolutionarily conserved and non-conserved retrovirus restriction activities of artiodactyl APOBEC3F proteins. *Nucleic Acids Res* 34(19), 5683-94.
- Katz, R.A. and Skalka, A.M. (1994) The retroviral enzymes. *Annu Rev Biochem* 63, 133-73.
- Katzourakis, A., Tristem, M., Pybus, O.G. and Gifford, R.J. (2007) Discovery and analysis of the first endogenous lentivirus. *Proc Natl Acad Sci U S A* 104(15), 6261-5.
- Kazazian, H.H., Jr. (2000) Genetics. L1 retrotransposons shape the mammalian genome. *Science* 289(5482), 1152-3.
- Kazazian, H.H., Jr. and Moran, J.V. (1998) The impact of L1 retrotransposons on the human genome. *Nat Genet* 19(1), 19-24.
- Khan, M.A., Goila-Gaur, R., Opi, S., Miyagi, E., Takeuchi, H., Kao, S. and Strebel, K. (2007) Analysis of the contribution of cellular and viral RNA to the packaging of APOBEC3G into HIV-1 virions. *Retrovirology* 4, 48.
- Khan, M.A., Kao, S., Miyagi, E., Takeuchi, H., Goila-Gaur, R., Opi, S., Gipson, C.L., Parslow, T.G., Ly, H. and Strebel, K. (2005) Viral RNA is required for the association of APOBEC3G with human immunodeficiency virus type 1 nucleoprotein complexes. *J Virol* 79(9), 5870-4.
- Kidd, J.M., Newman, T.L., Tuzun, E., Kaul, R. and Eichler, E.E. (2007) Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet* 3(4), e63.
- Kobayashi, M., Takaori-Kondo, A., Shindo, K., Abudu, A., Fukunaga, K. and Uchiyama, T. (2004) APOBEC3G targets specific virus species. *J Virol* 78(15), 8238-44.
- Krach, U., Fischer, N., Czauderna, F. and Tonjes, R.R. (2001) Comparison of replication-competent molecular clones of porcine endogenous retrovirus class A and class B derived from pig and human cells. *J Virol* 75(12), 5465-72.
- Kristbjornsdottir, H.B., Andresdottir, V., Svansson, V., Torsteinsdottir, S., Matthiasdottir, S. and Andresson, O.S. (2004) The vif gene of maedi-visna virus is essential for infectivity in vivo and in vitro. *Virology* 318(1), 350-9.
- LaRue, R.S., Andresdottir, V., Blanchard, Y., Conticello, S.G., Derse, D., Emerman, M., Greene, W.C., Jonsson, S.R., Landau, N.R., Lochelt, M., Malik, H.S., Malim, M.H., Munk, C., O'Brien, S.J., Pathak, V.K., Strebel, K., Wain-Hobson, S., Yu, X.F., Yuhki, N. and Harris, R.S. (2009) Guidelines for naming nonprimate APOBEC3 genes and proteins. *J Virol* 83(2), 494-7.
- LaRue, R.S., Jonsson, S.R., Silverstein, K.A., Lajoie, M., Bertrand, D., El-Mabrouk, N., Hotzel, I., Andresdottir, V., Smith, T.P. and Harris, R.S. (2008) The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. *BMC Mol Biol* 9, 104.
- Lee, Y.N. and Bieniasz, P.D. (2007) Reconstitution of an infectious human endogenous retrovirus. *PLoS Pathog* 3(1), e10.
- Liao, W., Hong, S.H., Chan, B.H., Rudolph, F.B., Clark, S.C. and Chan, L. (1999) APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. *Biochem Biophys Res Commun* 260(2), 398-404.

- Liddament, M.T., Brown, W.L., Schumacher, A.J. and Harris, R.S. (2004) APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. *Curr Biol* 14(15), 1385-91.
- Lin, F.H. and Thormar, H. (1970) Ribonucleic acid-dependent deoxyribonucleic acid polymerase in visna virus. *J Virol* 6(5), 702-4.
- List, J. and Haase, A.T. (1997) Integration of visna virus DNA occurs and may be necessary for productive infection. *Virology* 237(2), 189-97.
- Lockridge, K.M., Himathongkham, S., Sawai, E.T., Chienand, M. and Sparger, E.E. (1999) The feline immunodeficiency virus vif gene is required for productive infection of feline peripheral blood mononuclear cells and monocyte-derived macrophages. *Virology* 261(1), 25-30.
- Lower, R., Lower, J. and Kurth, R. (1996) The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc Natl Acad Sci U S A* 93(11), 5177-84.
- Luciw, P.A. and Leung, N.J. (1992) Mechanisms of retrovirus replication. In: J.A. Levy (Ed), *The Retroviridae*, pp. 159-298. Vol. 1. Plenum Press, New York.
- Luo, K., Liu, B., Xiao, Z., Yu, Y., Yu, X., Gorelick, R. and Yu, X.F. (2004) Amino-terminal region of the human immunodeficiency virus type 1 nucleocapsid is required for human APOBEC3G packaging. *J Virol* 78(21), 11841-52.
- Luo, K., Wang, T., Liu, B., Tian, C., Xiao, Z., Kappes, J. and Yu, X.F. (2007) Cytidine deaminases APOBEC3G and APOBEC3F interact with human immunodeficiency virus type 1 integrase and inhibit proviral DNA formation. *J Virol* 81(13), 7238-48.
- Lyall, J.W., Solanky, N. and Tiley, L.S. (2000) Restricted species tropism of maedi-visna virus strain EV-1 is not due to limited receptor distribution. *J Gen Virol* 81(Pt 12), 2919-27.
- MacDuff, D.A., Demorest, Z.L. and Harris, R.S. (2009) AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. *Nucleic Acids Res* 37(6), 1854-67.
- Madani, N. and Kabat, D. (1998) An endogenous inhibitor of human immunodeficiency virus in human lymphocytes is overcome by the viral Vif protein. *J Virol* 72(12), 10251-5.
- Magre, S., Takeuchi, Y. and Bartosch, B. (2003) Xenotransplantation and pig endogenous retroviruses. *Rev Med Virol* 13(5), 311-29.
- Malashkevich, V.N., Singh, M. and Kim, P.S. (2001) The trimer-of-hairpins motif in membrane fusion: Visna virus. *Proc Natl Acad Sci U S A* 98(15), 8502-6.
- Malim, M.H. and Emerman, M. (2008) HIV-1 accessory proteins--ensuring viral survival in a hostile environment. *Cell Host Microbe* 3(6), 388-98.
- Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L. and Trono, D. (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* 424, 99-103.
- Mariani, R., Chen, D., Schrofelbauer, B., Navarro, F., Konig, R., Bollman, B., Munk, C., Nymark-McMahon, H. and Landau, N.R. (2003) Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* 114(1), 21-31.
- Martin, U., Tacke, S.J., Simon, A.R., Schroder, C., Wiebe, K., Lapin, B., Haverich, A., Denner, J. and Steinhoff, G. (2002) Absence of PERV specific humoral immune response in baboons after transplantation of porcine cells or organs. *Transpl Int* 15(7), 361-8.
- Martin, U., Winkler, M.E., Id, M., Radeke, H., Arseniev, L., Takeuchi, Y., Simon, A.R., Patience, C., Haverich, A. and Steinhoff, G. (2000) Productive

- infection of primary human endothelial cells by pig endogenous retrovirus (PERV). *Xenotransplantation* 7(2), 138-42.
- Matsumoto, T., Marusawa, H., Endo, Y., Ueda, Y., Matsumoto, Y. and Chiba, T. (2006) Expression of APOBEC2 is transcriptionally regulated by NF-kappaB in human hepatocytes. *FEBS Lett* 580(3), 731-5.
- Mazarin, V., Gourdou, I., Querat, G., Sauze, N. and Vigne, R. (1988) Genetic structure and function of an early transcript of visna virus. *J Virol* 62(12), 4813-8.
- Mbisa, J.L., Barr, R., Thomas, J.A., Vandegraaff, N., Dorweiler, I.J., Svarovskaia, E.S., Brown, W.L., Mansky, L.M., Gorelick, R.J., Harris, R.S., Engelman, A. and Pathak, V.K. (2007) Human immunodeficiency virus type 1 cDNAs produced in the presence of APOBEC3G exhibit defects in plus-strand DNA transfer and integration. *J Virol* 81(13), 7099-110.
- Medstrand, P. and Mager, D.L. (1998) Human-specific integrations of the HERV-K endogenous retrovirus family. *J Virol* 72(12), 9782-7.
- Medstrand, P., van de Lagemaat, L.N. and Mager, D.L. (2002) Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res* 12(10), 1483-95.
- Meng, X.J., Purcell, R.H., Halbur, P.G., Lehman, J.R., Webb, D.M., Tsareva, T.S., Haynes, J.S., Thacker, B.J. and Emerson, S.U. (1997) A novel virus in swine is closely related to the human hepatitis E virus. *Proc Natl Acad Sci U S A* 94(18), 9860-5.
- Meyer, B.E., Meinkoth, J.L. and Malim, M.H. (1996) Nuclear transport of human immunodeficiency virus type 1, visna virus, and equine infectious anemia virus Rev proteins: identification of a family of transferable nuclear export signals. *J Virol* 70(4), 2350-9.
- Michaels, F.H., Hattori, N., Gallo, R.C. and Franchini, G. (1993) The human immunodeficiency virus type 1 (HIV-1) vif protein is located in the cytoplasm of infected cells and its effect on viral replication is equivalent in HIV-2. *AIDS Res Hum Retroviruses* 9(10), 1025-30.
- Mikkelsen, T.S., Wakefield, M.J., Aken, B., Amemiya, C.T., Chang, J.L., Duke, S., Garber, M., Gentles, A.J., Goodstadt, L., Heger, A., Jurka, J., Kamal, M., Mauceli, E., Searle, S.M., Sharpe, T., Baker, M.L., Batzer, M.A., Benos, P.V., Belov, K., Clamp, M., Cook, A., Cuff, J., Das, R., Davidow, L., Deakin, J.E., Fazzari, M.J., Glass, J.L., Grabherr, M., Greally, J.M., Gu, W., Hore, T.A., Huttley, G.A., Kleber, M., Jirtle, R.L., Koina, E., Lee, J.T., Mahony, S., Marra, M.A., Miller, R.D., Nicholls, R.D., Oda, M., Papenfuss, A.T., Parra, Z.E., Pollock, D.D., Ray, D.A., Schein, J.E., Speed, T.P., Thompson, K., VandeBerg, J.L., Wade, C.M., Walker, J.A., Waters, P.D., Webber, C., Weidman, J.R., Xie, X., Zody, M.C., Graves, J.A., Ponting, C.P., Breen, M., Samollow, P.B., Lander, E.S. and Lindblad-Toh, K. (2007) Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447(7141), 167-77.
- Miyagi, E., Opi, S., Takeuchi, H., Khan, M., Goila-Gaur, R., Kao, S. and Strebel, K. (2007) Enzymatically active APOBEC3G is required for efficient inhibition of human immunodeficiency virus type 1. *J Virol* 81(24), 13346-53.
- Monie, T.P., Greatorex, J.S., Maynard-Smith, L., Hook, B.D., Bishop, N., Beales, L.P. and Lever, A.M. (2005) Identification and visualization of the dimerization initiation site of the prototype lentivirus, maedi visna virus: a

- potential GACG tetraloop displays structural homology with the alpha- and gamma-retroviruses. *Biochemistry* 44(1), 294-302.
- Munk, C., Beck, T., Zielonka, J., Hotz-Wagenblatt, A., Chareza, S., Battenberg, M., Thielebein, J., Cichutek, K., Bravo, I.G., O'Brien, S.J., Lochelt, M. and Yuhki, N. (2008) Functions, structure, and read-through alternative splicing of feline APOBEC3 genes. *Genome Biol* 9(3), R48.
- Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y. and Honjo, T. (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102(5), 553-63.
- Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O. and Honjo, T. (1999) Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem* 274(26), 18470-6.
- Nathans, R., Cao, H., Sharova, N., Ali, A., Sharkey, M., Stranska, R., Stevenson, M. and Rana, T.M. (2008) Small-molecule inhibition of HIV-1 Vif. *Nat Biotechnol* 26(10), 1187-92.
- Navarro, R.E. and Blackwell, T.K. (2005) Requirement for P granules and meiosis for accumulation of the germline RNA helicase CGH-1. *Genesis* 42(3), 172-80.
- Neil, S.J., Zang, T. and Bieniasz, P.D. (2008) Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* 451(7177), 425-30.
- Newman, E.N., Holmes, R.K., Craig, H.M., Klein, K.C., Lingappa, J.R., Malim, M.H. and Sheehy, A.M. (2005) Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr Biol* 15(2), 166-70.
- Niebert, M., Rogel-Gaillard, C., Chardon, P. and Tonjes, R.R. (2002) Characterization of chromosomally assigned replication-competent gamma porcine endogenous retroviruses derived from a large white pig and expression in human cells. *J Virol* 76(6), 2714-20.
- Oberste, M.S. and Gonda, M.A. (1992) Conservation of amino-acid sequence motifs in lentivirus Vif proteins. *Virus Genes* 6(1), 95-102.
- Ochsenbauer, C., Wilk, T. and Bosch, V. (1997) Analysis of vif-defective human immunodeficiency virus type 1 (HIV-1) virions synthesized in 'non-permissive' T lymphoid cells stably infected with selectable HIV-1. *J Gen Virol* 78(Pt 3), 627-35.
- OhAinle, M., Kerns, J.A., Li, M.M., Malik, H.S. and Emerman, M. (2008) Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe* 4(3), 249-59.
- Oldmixon, B.A., Wood, J.C., Ericsson, T.A., Wilson, C.A., White-Scharf, M.E., Andersson, G., Greenstein, J.L., Schuurman, H.J. and Patience, C. (2002) Porcine endogenous retrovirus transmission characteristics of an inbred herd of miniature swine. *J Virol* 76(6), 3045-8.
- Ono, M., Kawakami, M. and Takezawa, T. (1987) A novel human nonviral retroposon derived from an endogenous retrovirus. *Nucleic Acids Res* 15(21), 8725-37.
- Oskarsson, T., Hreggvidsdottir, H.S., Agnarsdottir, G., Matthiasdottir, S., Ogmundsdottir, M.H., Jonsson, S.R., Georgsson, G., Ingvarsson, S., Andresson, O.S. and Andresdottir, V. (2007) Duplicated sequence motif in the long terminal repeat of maedi-visna virus extends cell tropism and is associated with neurovirulence. *J Virol* 81(8), 4052-7.

- Ostertag, E.M., Goodier, J.L., Zhang, Y. and Kazazian, H.H., Jr. (2003) SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am J Hum Genet* 73(6), 1444-51.
- Paillart, J.C., Shehu-Xhilaga, M., Marquet, R. and Mak, J. (2004) Dimerization of retroviral RNA genomes: an inseparable pair. *Nat Rev Microbiol* 2(6), 461-72.
- Palsson, P.A. (1976) Maedi and visna in sheep. *Front Biol* 44, 17-43.
- Panganiban, A.T. and Fiore, D. (1988) Ordered interstrand and intrastrand DNA transfer during reverse transcription. *Science* 241(4869), 1064-9.
- Park, I.W., Myrick, K. and Sodroski, J. (1994) Effects of vif mutations on cell-free infectivity and replication of simian immunodeficiency virus. *J Acquir Immune Defic Syndr* 7(12), 1228-36.
- Patience, C., Takeuchi, Y. and Weiss, R.A. (1997) Infection of human cells by an endogenous retrovirus of pigs. *Nat Med* 3(3), 282-6.
- Paton, N.I., Leo, Y.S., Zaki, S.R., Auchus, A.P., Lee, K.E., Ling, A.E., Chew, S.K., Ang, B., Rollin, P.E., Umaphathi, T., Sng, I., Lee, C.C., Lim, E. and Ksiazek, T.G. (1999) Outbreak of Nipah-virus infection among abattoir workers in Singapore. *Lancet* 354(9186), 1253-6.
- Pepin, M., Vitu, C., Russo, P., Mornex, J.F. and Peterhans, E. (1998) Maedi-visna virus infection in sheep: a review. *Vet Res* 29(3-4), 341-67.
- Petersen-Mahrt, S.K., Harris, R.S. and Neuberger, M.S. (2002) AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* 418(6893), 99-103.
- Petursson, G., Matthiasdottir, S., Svansson, V., Andresdottir, V., Georgsson, G., Martin, A.H., Agnarsdottir, G., Gisladdottir, E., Arnadottir, S., Hognadottir, S., Jonsson, S.R., Andresson, O.S. and Torsteinsdottir, S. (2005) Mucosal vaccination with an attenuated maedi-visna virus clone. *Vaccine* 23(24), 3223-8.
- Petursson, G., Nathanson, N., Georgsson, G., Panitch, H. and Palsson, P.A. (1976) Pathogenesis of visna. I. Sequential virologic, serologic, and pathologic studies. *Lab Invest* 35(4), 402-12.
- Petursson, G., Turelli, P., Matthiasdottir, S., Georgsson, G., Andresson, O.S., Torsteinsdottir, S., Vigne, R., Andresdottir, V., Gunnarsson, E., Agnarsdottir, G. and Querat, G. (1998) Visna virus dUTPase is dispensable for neuropathogenicity. *J Virol* 72(2), 1657-61.
- Prochnow, C., Bransteitter, R., Klein, M.G., Goodman, M.F. and Chen, X.S. (2007) The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* 445(7126), 447-51.
- Prusiner, S.B. (1982) Research on scrapie. *Lancet* 2(8296), 494-5.
- Prusiner, S.B. and Scott, M.R. (1997) Genetics of prions. *Annu Rev Genet* 31, 139-75.
- Rabson, A.B. and Graves, B.J. (1997) Synthesis and processing of viral RNA. In: J.M. Coffin, S.H. Hughes and H.E. Varmus (Eds), *Retroviruses*, pp. 205-261. Cold Spring Harbor Laboratory Press.
- Reddy, T.R., Kraus, G., Suhasini, M., Leavitt, M.C. and Wong-Staal, F. (1995) Identification and mapping of inhibitory sequences in the human immunodeficiency virus type 2 vif gene. *J Virol* 69(8), 5167-70.
- Rogozin, I.B., Basu, M.K., Jordan, I.K., Pavlov, Y.I. and Koonin, E.V. (2005) APOBEC4, a new member of the AID/APOBEC family of polynucleotide

- (deoxy)cytidine deaminases predicted by computational analysis. *Cell Cycle* 4(9), 1281-5.
- Rous, P. (1911) Transmission of a malignant new growth by means of a cell-free filtrate. *JAMA* 56, 198.
- Russell, R.A. and Pathak, V.K. (2007) Identification of two distinct human immunodeficiency virus type 1 Vif determinants critical for interactions with human APOBEC3G and APOBEC3F. *J Virol* 81(15), 8201-10.
- Russell, R.A., Wiegand, H.L., Moore, M.D., Schafer, A., McClure, M.O. and Cullen, B.R. (2005) Foamy virus Bet proteins function as novel inhibitors of the APOBEC3 family of innate antiretroviral defense factors. *J Virol* 79(14), 8724-31.
- Santos, N., Lima, R.C., Nozawa, C.M., Linhares, R.E. and Gouvea, V. (1999) Detection of porcine rotavirus type G9 and of a mixture of types G1 and G5 associated with Wa-like VP4 specificity: evidence for natural human-porcine genetic reassortment. *J Clin Microbiol* 37(8), 2734-6.
- Sargan, D.R. and Bennet, I.D. (1989) A transcriptional map of visna virus: definition of the second intron structure suggests a rev-like gene product. *J Gen Virol* 70 ( Pt 8), 1995-2006.
- Sawyer, S.L., Emerman, M. and Malik, H.S. (2004) Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol* 2(9), E275.
- Schafer, A., Bogerd, H.P. and Cullen, B.R. (2004) Specific packaging of APOBEC3G into HIV-1 virions is mediated by the nucleocapsid domain of the gag polyprotein precursor. *Virology* 328(2), 163-8.
- Schrofelbauer, B., Chen, D. and Landau, N.R. (2004) A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). *Proc Natl Acad Sci U S A* 101(11), 3927-32.
- Schumacher, A.J., Hache, G., Macduff, D.A., Brown, W.L. and Harris, R.S. (2008) The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction. *J Virol* 82(6), 2652-60.
- Schumacher, A.J., Nissley, D.V. and Harris, R.S. (2005) APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. *Proc Natl Acad Sci U S A* 102(28), 9854-9.
- Schuurman, H.J., Cheng, J. and Lam, T. (2003) Pathology of xenograft rejection: a commentary. *Xenotransplantation* 10(4), 293-9.
- Shacklett, B.L. and Luciw, P.A. (1994) Analysis of the vif gene of feline immunodeficiency virus. *Virology* 204(2), 860-7.
- Sheehy, A.M., Gaddis, N.C., Choi, J.D. and Malim, M.H. (2002) Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 418(6898), 646-50.
- Sheehy, A.M., Gaddis, N.C. and Malim, M.H. (2003) The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nature Medicine* 9(11), 1404-7.
- Shindo, K., Takaori-Kondo, A., Kobayashi, M., Abudu, A., Fukunaga, K. and Uchiyama, T. (2003) The enzymatic activity of CEM15/Apobec-3G is essential for the regulation of the infectivity of HIV-1 virion but not a sole determinant of its antiviral activity. *J Biol Chem* 278(45), 44412-6.

- Sigurdsson, B. (1954a) Observations on three slow infections of sheep (1). Maedi, a slow progressive pneumonia of sheep: An epizootological and pathological study. *British Veterinary Journal* 110, 255-270.
- Sigurdsson, B. (1954b) Observations on three slow infections of sheep (3). Rida, a chronic encephalitis of sheep. With general remarks on infections which develop slowly and some of their special characters. *British Veterinary Journal* 110, 341-354.
- Sigurdsson, B., Palsson, P. and Grimsson, H. (1957) Visna, a demyelinating transmissible disease of sheep. *J Neuropathol Exp Neurol* 16(3), 389-403.
- Sigurdsson, B. and Palsson, P.A. (1958) Visna of sheep. A slow demyelinating infection. *Br J Exp Pathol* 39, 519-528.
- Sigurdsson, B., Palsson, P.A. and Tryggvaddottir, A. (1953) Transmission experiments with maedi. *J Infect Dis* 93(2), 166-75.
- Sigurdsson, B., Thormar, H. and Palsson, P.A. (1960) Cultivation of visna virus in tissue culture. *Archiv fur die gesamte Virusforschung* 10, 368-380.
- Simon, J.H., Fouchier, R.A., Southerling, T.E., Guerra, C.B., Grant, C.K. and Malim, M.H. (1997) The Vif and Gag proteins of human immunodeficiency virus type 1 colocalize in infected human T cells. *J Virol* 71(7), 5259-67.
- Simon, J.H., Gaddis, N.C., Fouchier, R.A. and Malim, M.H. (1998a) Evidence for a newly discovered cellular anti-HIV-1 phenotype. *Nat Med* 4(12), 1397-400.
- Simon, J.H., Miller, D.L., Fouchier, R.A., Soares, M.A., Peden, K.W. and Malim, M.H. (1998b) The regulation of primate immunodeficiency virus infectivity by Vif is cell species restricted: a role for Vif in determining virus host range and cross-species transmission. *Embo J* 17(5), 1259-67.
- Skraban, R., Matthiasdottir, S., Torsteinsdottir, S., Agnarsdottir, G., Gudmundsson, B., Georgsson, G., Meloen, R.H., Andresson, O.S., Staskus, K.A., Thormar, H. and Andresdottir, V. (1999) Naturally occurring mutations within 39 amino acids in the envelope glycoprotein of maedi-visna virus alter the neutralization phenotype. *J Virol* 73(10), 8064-72.
- Smith, T.P., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Pertea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keele, J.W. (2001) Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle. *Genome Res* 11(4), 626-30.
- Soin, B., Ostlie, D., Cozzi, E., Smith, K.G., Bradley, J.R., Vial, C., Masroor, S., Lancaster, R., White, D.J. and Friend, P.J. (2000) Growth of porcine kidneys in their native and xenograft environment. *Xenotransplantation* 7(2), 96-100.
- Sonigo, P., Alizon, M., Staskus, K., Klatzmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, P., Haase, A. and Wain-Hobson, S. (1985) Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. *Cell* 42(1), 369-82.
- Sova, P. and Volsky, D.J. (1993) Efficiency of viral DNA synthesis during infection of permissive and nonpermissive cells with vif-negative human immunodeficiency virus type 1. *J Virol* 67(10), 6322-6.
- Stanley, B.J., Ehrlich, E.S., Short, L., Yu, Y., Xiao, Z., Yu, X.F. and Xiong, Y. (2008) Structural insight into the human immunodeficiency virus Vif SOCS box and its role in human E3 ubiquitin ligase assembly. *J Virol* 82(17), 8656-63.
- Staskus, K.A., Retzel, E.F., Lewis, E.D., Silsby, J.L., St Cyr, S., Rank, J.M., Wietgreffe, S.W., Haase, A.T., Cook, R. and Fast, D. (1991) Isolation of

- replication-competent molecular clones of visna virus. *Virology* 181(1), 228-40.
- Stenglein, M.D. and Harris, R.S. (2006) APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J Biol Chem* 281(25), 16837-41.
- Stoye, J.P. (2006) Koala retrovirus: a genome invasion in real time. *Genome Biol* 7(11), 241.
- Straub, O.C. (2004) Maedi-Visna virus infection in sheep. History and present knowledge. *Comp Immunol Microbiol Infect Dis* 27(1), 1-5.
- Strebel, K., Daugherty, D., Clouse, K., Cohen, D., Folks, T. and Martin, M.A. (1987) The HIV 'A' (sor) gene product is essential for virus infectivity. *Nature* 328(6132), 728-30.
- Svarovskaia, E.S., Xu, H., Mbisa, J.L., Barr, R., Gorelick, R.J., Ono, A., Freed, E.O., Hu, W.S. and Pathak, V.K. (2004) Human apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G) is incorporated into HIV-1 virions through interactions with viral and nonviral RNAs. *J Biol Chem* 279(34), 35822-8.
- Swanstrom, R. and Wills, J.W. (1997) Synthesis, assembly, and processing of viral proteins. In: J.M. Coffin, S.H. Hughes and H.E. Varmus (Eds), *Retroviruses*, pp. 263-334. Cold Spring Harbor Laboratory Press.
- Swindle, M.M. (1998) Defining appropriate health status and management programs for specific-pathogen-free swine for xenotransplantation. *Ann N Y Acad Sci* 862, 111-20.
- Tarlinton, R., Meers, J., Hanger, J. and Young, P. (2005) Real-time reverse transcriptase PCR for the endogenous koala retrovirus reveals an association between plasma viral load and neoplastic disease in koalas. *J Gen Virol* 86(Pt 3), 783-7.
- Tarlinton, R., Meers, J. and Young, P. (2008) Biology and evolution of the endogenous koala retrovirus. *Cell Mol Life Sci* 65(21), 3413-21.
- Telesnitsky, A. and Goff, S.P. (1997) Reverse transcriptase and the generation of retroviral DNA. *Retroviruses*, edited by J.M. Coffin, Huges, S.H., Varmus, H.E. Cold Spring Harbor Laboratory Press.
- Temin, H.M. and Mizutani, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 226(5252), 1211-3.
- Teng, B., Burant, C.F. and Davidson, N.O. (1993) Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* 260(5115), 1816-9.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22), 4673-80.
- Thormar, H. (2005) Maedi-visna virus and its relationship to human immunodeficiency virus. *AIDS Rev* 7(4), 233-45.
- Tiley, L.S., Brown, P.H., Le, S.Y., Maizel, J.V., Clements, J.E. and Cullen, B.R. (1990) Visna virus encodes a post-transcriptional regulator of viral structural gene expression. *Proc Natl Acad Sci U S A* 87(19), 7497-501.
- Tiley, L.S. and Cullen, B.R. (1992) Structural and functional analysis of the visna virus Rev-response element. *J Virol* 66(6), 3609-15.

- Tiley, L.S., Malim, M.H. and Cullen, B.R. (1991) Conserved functional organization of the human immunodeficiency virus type 1 and visna virus Rev proteins. *J Virol* 65(7), 3877-81.
- Todaro, G.J., Benveniste, R.E., Lieber, M.M. and Sherr, C.J. (1974) Characterization of a type C virus released from the porcine cell line PK(15). *Virology* 58(1), 65-74.
- Tomonaga, K., Norimine, J., Shin, Y.S., Fukasawa, M., Miyazawa, T., Adachi, A., Toyosaki, T., Kawaguchi, Y., Kai, C. and Mikami, T. (1992) Identification of a feline immunodeficiency virus gene which is essential for cell-free virus infectivity. *J Virol* 66(10), 6181-5.
- Torsteinsdottir, S., Agnarsdottir, G., Matthiasdottir, S., Rafnar, B., Andresdottir, V., Andresson, O.S., Staskus, K., Petursson, G., Palsson, P.A. and Georgsson, G. (1997) In vivo and in vitro infection with two different molecular clones of visna virus. *Virology* 229(2), 370-80.
- Torsteinsdottir, S., Matthiasdottir, S., Vidarsdottir, N., Svansson, V. and Petursson, G. (2003) Intratracheal inoculation as an efficient route of experimental infection with maedi-visna virus. *Res Vet Sci* 75(3), 245-7.
- Tseng, Y.L., Dor, F.J., Kuwaki, K., Ryan, D., Wood, J., Denaro, M., Giovino, M., Yamada, K., Hawley, R., Patience, C., Schuurman, H.J., Awwad, M., Sachs, D.H. and Cooper, D.K. (2004) Bone marrow transplantation from alpha1,3-galactosyltransferase gene-knockout pigs in baboons. *Xenotransplantation* 11(4), 361-70.
- Ulrich, S., Goltz, M. and Ehlers, B. (1999) Characterization of the DNA polymerase loci of the novel porcine lymphotropic herpesviruses 1 and 2 in domestic and feral pigs. *J Gen Virol* 80 ( Pt 12), 3199-205.
- van de Lagemaat, L.N., Landry, J.R., Mager, D.L. and Medstrand, P. (2003) Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet* 19(10), 530-6.
- Vigne, R., Barban, V., Quérat, G., Mazarin, V., Gourdou, I. and Sauze, N. (1987) Transcription of visna virus during its lytic cycle: evidence for a sequential early and late gene expression. *Virology* 161(1), 218-27.
- Vigne, R., Neuveut, C., Sire, J., Philippon, V., Filippi, P., Gambarelli, D., Clements, J.E., Hayman, M., Arbuthnott, G. and Harkiss, G. (1994) Involvement of viral regulatory gene products in the pathogenesis of lentivirus infections. *Ann N Y Acad Sci* 724, 107-24.
- Villet, S., Bouzar, B.A., Morin, T., Verdier, G., Legras, C. and Chebloune, Y. (2003) Maedi-visna virus and caprine arthritis encephalitis virus genomes encode a Vpr-like but no Tat protein. *J Virol* 77(17), 9632-8.
- Vogt, V.M. (1997) Retroviral virions and genomes. In: J.M. Coffin, S.H. Hughes and H.E. Varmus (Eds), *Retroviruses*, pp. 27-70. Cold Spring Harbor Laboratory Press.
- Volsky, D.J., Potash, M.J., Simm, M., Sova, P., Ma, X.Y., Chao, W. and Shahabuddin, M. (1995) The human immunodeficiency virus type 1 vif gene: the road from an accessory to an essential role in human immunodeficiency virus type 1 replication. *Curr Top Microbiol Immunol* 193, 157-68.
- von Schwedler, U., Song, J., Aiken, C. and Trono, D. (1993) Vif is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells. *J Virol* 67(8), 4945-55.

- Wang, T., Tian, C., Zhang, W., Luo, K., Sarkis, P.T., Yu, L., Liu, B., Yu, Y. and Yu, X.F. (2007) 7SL RNA mediates virion packaging of the antiviral cytidine deaminase APOBEC3G. *J Virol* 81(23), 13112-24.
- Warren, W.C., Hillier, L.W., Marshall Graves, J.A., Birney, E., Ponting, C.P., Grutzner, F., Belov, K., Miller, W., Clarke, L., Chinwalla, A.T., Yang, S.P., Heger, A., Locke, D.P., Miethke, P., Waters, P.D., Veyrunes, F., Fulton, L., Fulton, B., Graves, T., Wallis, J., Puente, X.S., Lopez-Otin, C., Ordenez, G.R., Eichler, E.E., Chen, L., Cheng, Z., Deakin, J.E., Alsop, A., Thompson, K., Kirby, P., Papenfuss, A.T., Wakefield, M.J., Olender, T., Lancet, D., Huttley, G.A., Smit, A.F., Pask, A., Temple-Smith, P., Batzer, M.A., Walker, J.A., Konkel, M.K., Harris, R.S., Whittington, C.M., Wong, E.S., Gemmell, N.J., Buschiazzi, E., Vargas Jentzsch, I.M., Merkel, A., Schmitz, J., Zemmann, A., Churakov, G., Kriegs, J.O., Brosius, J., Murchison, E.P., Sachidanandam, R., Smith, C., Hannon, G.J., Tsend-Ayush, E., McMillan, D., Attenborough, R., Rens, W., Ferguson-Smith, M., Lefevre, C.M., Sharp, J.A., Nicholas, K.R., Ray, D.A., Kube, M., Reinhardt, R., Pringle, T.H., Taylor, J., Jones, R.C., Nixon, B., Dacheux, J.L., Niwa, H., Sekita, Y., Huang, X., Stark, A., Kheradpour, P., Kellis, M., Flicek, P., Chen, Y., Webber, C., Hardison, R., Nelson, J., Hallsworth-Pepin, K., Delehaunty, K., Markovic, C., Minx, P., Feng, Y., Kremitzki, C., Mitreva, M., Glasscock, J., Wylie, T., Wohldmann, P., Thiru, P., Nhan, M.N., Pohl, C.S., Smith, S.M., Hou, S., Nefedov, M., et al. (2008) Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453(7192), 175-83.
- Wedekind, J.E., Dance, G.S., Sowden, M.P. and Smith, H.C. (2003) Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. *Trends Genet* 19(4), 207-16.
- Wei, X., Decker, J.M., Wang, S., Hui, H., Kappes, J.C., Wu, X., Salazar-Gonzalez, J.F., Salazar, M.G., Kilby, J.M., Saag, M.S., Komarova, N.L., Nowak, M.A., Hahn, B.H., Kwong, P.D. and Shaw, G.M. (2003) Antibody neutralization and escape by HIV-1. *Nature* 422(6929), 307-12.
- Weiss, R.A. (1998) Transgenic pigs and virus adaptation. *Nature* 391(6665), 327-8.
- Weiss, R.A. (2006) The discovery of endogenous retroviruses. *Retrovirology* 3, 67.
- Wells, D.L., Hopfensperger, D.J., Arden, N.H., Harmon, M.W., Davis, J.P., Tipple, M.A. and Schonberger, L.B. (1991) Swine influenza virus infections. Transmission from ill pigs to humans at a Wisconsin agricultural fair and subsequent probable person-to-person transmission. *JAMA* 265(4), 478-81.
- Wiegand, H.L., Doehle, B.P., Bogerd, H.P. and Cullen, B.R. (2004) A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. *Embo J* 23(12), 2451-8.
- Williams, T.P., Kasorndorkbua, C., Halbur, P.G., Haqshenas, G., Guenette, D.K., Toth, T.E. and Meng, X.J. (2001) Evidence of extrahepatic sites of replication of the hepatitis E virus in a swine model. *J Clin Microbiol* 39(9), 3040-6.
- Wilson, C.A., Wong, S., VanBrocklin, M. and Federspiel, M.J. (2000) Extended analysis of the in vitro tropism of porcine endogenous retrovirus. *J Virol* 74(1), 49-56.
- Wood, A., Webb, B.L., Bartosch, B., Schaller, T., Takeuchi, Y. and Towers, G.J. (2009) Porcine endogenous retroviruses PERV A and A/C recombinant are insensitive to a range of divergent mammalian TRIM5alpha proteins including human TRIM5alpha. *J Gen Virol* 90(Pt 3), 702-9.

- Yamanaka, S., Balestra, M.E., Ferrell, L.D., Fan, J., Arnold, K.S., Taylor, S., Taylor, J.M. and Innerarity, T.L. (1995) Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proc Natl Acad Sci U S A* 92(18), 8483-7.
- Yu, X., Yu, Y., Liu, B., Luo, K., Kong, W., Mao, P. and Yu, X.F. (2003) Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* 302(5647), 1056-60.
- Yu, Y., Xiao, Z., Ehrlich, E.S., Yu, X. and Yu, X.F. (2004) Selective assembly of HIV-1 Vif-Cul5-ElonginB-ElonginC E3 ubiquitin ligase complex through a novel SOCS box and upstream cysteines. *Genes Dev* 18(23), 2867-72.
- Zennou, V., Perez-Caballero, D., Gottlinger, H. and Bieniasz, P.D. (2004) APOBEC3G incorporation into human immunodeficiency virus type 1 particles. *J Virol* 78(21), 12058-61.
- Zhang, H., Yang, B., Pomerantz, R.J., Zhang, C., Arunachalam, S.C. and Gao, L. (2003) The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* 424, 94-98.
- Zhang, J. and Webb, D.M. (2004) Rapid evolution of primate antiviral enzyme APOBEC3G. *Hum Mol Genet* 13(16), 1785-91.
- Zhang, K.L., Mangeat, B., Ortiz, M., Zoete, V., Trono, D., Telenti, A. and Michielin, O. (2007) Model structure of human APOBEC3G. *PLoS ONE* 2(4), e378.
- Zhang, W., Chen, G., Niewiadomska, A.M., Xu, R. and Yu, X.F. (2008) Distinct determinants in HIV-1 Vif and human APOBEC3 proteins are required for the suppression of diverse host anti-viral proteins. *PLoS ONE* 3(12), e3963.



**7. PAPERS I – IV**





# Paper I

## Simultaneous Mutations in CA and Vif of Maedi-Visna Virus Cause Attenuated Replication in Macrophages and Reduced Infectivity In Vivo

Bjarki Gudmundsson,<sup>1</sup> Stefán Ragnar Jónsson,<sup>1,2</sup> Oddur Ólafsson,<sup>1</sup> Gudrún Agnarsdóttir,<sup>1</sup>  
Sigríður Matthíasdóttir,<sup>1</sup> Gudmundur Georgsson,<sup>1</sup> Sigurbjörg Torsteinsdóttir,<sup>1</sup>  
Vilhjálmur Svansson,<sup>1</sup> Helga Bryndís Kristbjörnsdóttir,<sup>1</sup>  
Sigríður Rut Franzdóttir,<sup>1</sup> Ólafur S. Andrésson,<sup>1</sup>  
and Valgerður Andréssdóttir<sup>1\*</sup>

*Institute for Experimental Pathology, University of Iceland, Keldur, Reykjavík, Iceland,<sup>1</sup> and Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, Minnesota<sup>2</sup>*

Received 8 July 2005/Accepted 20 September 2005

**Maedi-visna virus (MVV) is a lentivirus of sheep sharing several key features with the primate lentiviruses. The virus causes slowly progressive diseases, mainly in the lungs and the central nervous system of sheep. Here, we investigate the molecular basis for the differential growth phenotypes of two MVV isolates. One of the isolates, KV1772, replicates well in a number of cell lines and is highly pathogenic in sheep. The second isolate, KS1, no longer grows on macrophages or causes disease. The two virus isolates differ by 129 nucleotide substitutions and two deletions of 3 and 15 nucleotides in the *env* gene. To determine the molecular nature of the lesions responsible for the restrictive growth phenotype, chimeric viruses were constructed and used to map the phenotype. An L120R mutation in the CA domain, together with a P205S mutation in Vif (but neither alone), could fully convert KV1772 to the restrictive growth phenotype. These results suggest a functional interaction between CA and Vif in MVV replication, a property that may relate to the innate antiretroviral defense mechanisms in sheep.**

Maedi-visna virus (MVV) is a member of the lentivirus subfamily of retroviruses, causing encephalitis (visna), pneumonia (maedi), mastitis, and arthritis in sheep (26, 27). The primary target cells of MVV infection are cells of the monocyte/macrophage lineage, and virus expression is activated upon macrophage maturation (6, 11, 22). However, various cell types are permissive for infection by certain MVV strains in vitro. Lentiviruses in general exhibit a highly restricted cell and species tropism. This restriction is largely determined by host factors that either are required for virus replication or have antiviral activity.

Recently, two classes of lentiviral inhibitors were identified. One class is exemplified by APOBEC3G, a cytosine deaminase that induces C-to-U lesions in the viral cDNA minus strand during reverse transcription. This leads either to plus-strand G→A hypermutation in integrated virus DNA (C→T on the minus strand) or possibly also to the degradation of the retroviral cDNA prior to integration (14, 19, 21, 38). The *vif* gene, which is present in all lentiviruses except equine infectious anemia virus, counteracts APOBEC3G by causing its degradation. Vif is essential for virus infectivity in vivo and for replication in primary cells in vitro (5, 12, 13, 18, 20, 31).

We have shown that MVV lacking Vif is noninfectious in sheep and that replication is cell type dependent. Thus, Vif-deficient MVV replicates poorly in macrophages and sheep

choroid plexus (SCP) cells, whereas fetal ovine synovial (FOS) cells appear semipermissive. We have also found an elevated frequency of G→A mutations in Vif-deficient MVV, indicating a function for MVV Vif that is similar to that of human immunodeficiency virus type 1 (HIV-1) (18).

The capsid of retroviruses has been shown to be a target for cellular restriction factors in various virus and cell systems. The first of these to be characterized was Fv1, a gene product of an endogenous retrovirus in the mouse genome that blocks certain strains of Friend leukemia virus. The amino acid sequence of the capsid determines whether the virus is sensitive to Fv1 or not (4). A host factor, Lv1, restricting HIV-1 in many monkey cells has recently been identified as tripartite motif protein 5α (TRIM5α), a protein of unknown function (32). A similar restriction has been found in human cells that are resistant to certain murine leukemia virus strains, and the capsid is targeted here as well (15, 35). The restriction factor, Ref1, has also been shown to be a TRIM5α variant (17).

In this study, we examined the different cell tropism and pathogenicity of two molecular clones of MVV (2, 33). One of the clones (KS1) replicates well in SCP cells, but replicates only to low titers in macrophages, and is nonpathogenic in sheep. The other clone (KV1772) replicates to high titers in SCP cells and macrophages and is highly pathogenic in sheep. Both clones have their origin in a transmission experiment where virus was passaged serially through tissue culture and sheep (29), and they differ by 129 nucleotides. By constructing recombinant viruses and a series of amino acid substitution mutations, we mapped the attenuated phenotype of KS1 to two point mutations, one in the CA domain of *gag* and the other in *vif*.

\* Corresponding author. Mailing address: Institute for Experimental Pathology, University of Iceland, Keldur v/Vesturlandsveg, 112 Reykjavík, Iceland. Phone: 354-5855100. Fax: 354-5673979. E-mail: valand@hi.is.

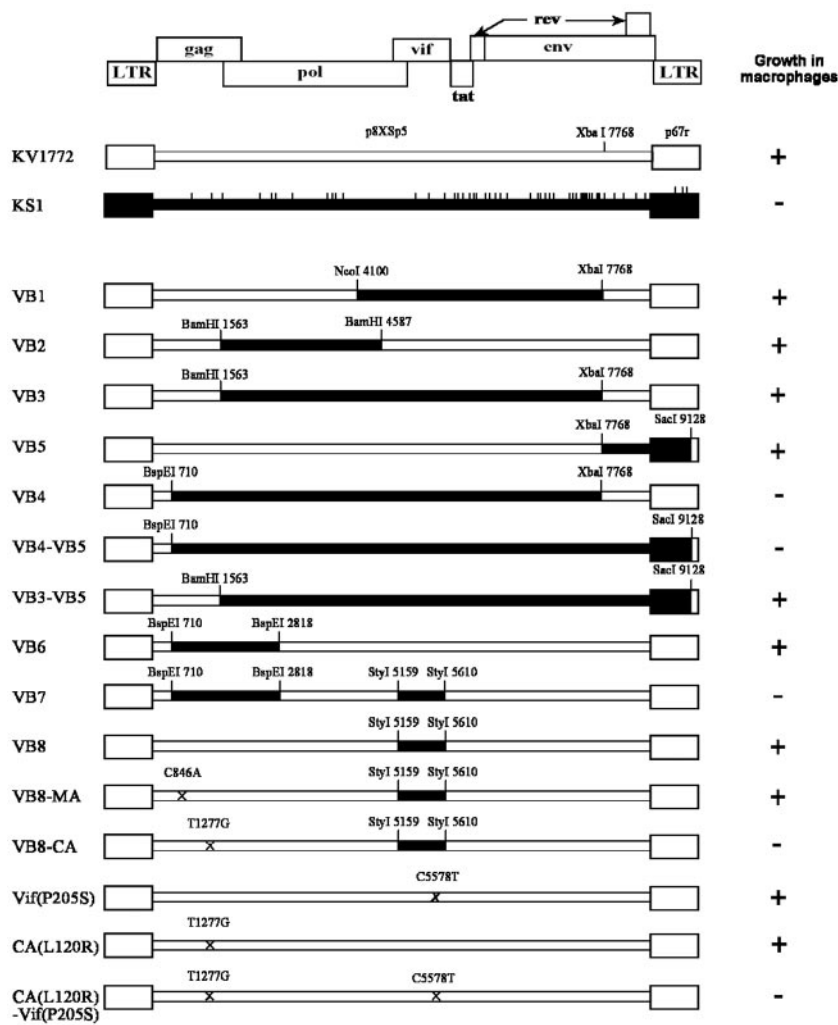


FIG. 1. Diagrammatic representation of recombinant MVV derived from the molecular clones KV1772 and KS1. The vertical bars indicate the amino acid differences between the two clones. The restriction sites used to construct the viruses are shown. The ability of each virus to grow in macrophages is shown on the right.

**MATERIALS AND METHODS**

**Virus and cells.** The molecularly cloned viruses KV1772 (formerly KV1772kv72/67) and KS1 (formerly LV1-1KS1) have been described previously (2, 29, 30). SCP cells established as described previously (7, 28) were grown at 37°C in a humidified atmosphere of 5% CO<sub>2</sub> in Dulbecco's modified Eagle medium (Gibco) supplemented with 2 mM glutamine, 100 IU of penicillin per ml, 100 IU of streptomycin per ml, and either 10% lamb serum (growth medium) or 1% lamb serum (maintenance medium). Macrophage cultures were established as described previously (29).

Transfections were performed by using FOS cells. DNA was transfected with Lipofectamine as specified by the manufacturer (Life Sciences, Inc.). Transfected FOS cells were passaged and incubated in maintenance medium until syncytia appeared (5 to 8 days). Supernatants from transfected cells were also tested for the presence of reverse transcriptase (RT) activity before passage into SCP cells.

**RT assay.** Viral particles from 0.5 ml of cell-free supernatants from infected cells were pelleted at 14,000 rpm for 1 h in a microfuge. RT activity was determined as described previously (1).

**Real-time PCR assay.** Viral particles from 200 µl of cell-free supernatants from infected cells were pelleted at 14,000 rpm for 1 h in a microfuge. The pellet was dissolved in 10 µl TNE (10 mM Tris, pH 7.5; 100 mM NaCl; 1 mM EDTA) with 0.1% Triton X-100. This lysate was used for generating cDNA with ReverTaid M-murine leukemia virus reverse transcriptase (Fermentas) and a primer

from the gag gene (V-1818 [5'-CGGGGTACCTTACAACATAGGGGGCGCG G-3']). Real-time PCR was carried out with a DNA Engine Opticon System (MJ Research) in a final volume of 20 µl. The primers and TaqMan probe were as follows: forward primer, V1636 (5'-TAAATCAAAAGTGTATAATTGTGGG A-3'), reverse primer, V-1719 (5'-TCCCACAATGATGGCATATTATTTC-3'); TaqMan probe, V1665TaqMan (5'-FAM-CCAGGACATCTCGCAAGACAGT GTAGACA-Black Hole Quencher-1-3'). (FAM is 6-carboxyfluorescein.) Calibration curves were derived by running 10-fold dilutions of specific cDNA over a range of 60 to 6 × 10<sup>7</sup> copies. Each assay included duplicate wells for each dilution of calibration DNA and for each cDNA sample.

**Construction of recombinant clones.** The molecular clone KV1772 is contained in two plasmids as described previously (29). The restriction enzymes used for cloning are indicated in Fig. 1. Mutants were generated by PCR-mediated site-directed mutagenesis, and the mutations were subsequently confirmed by sequencing.

**Experimental infection of sheep.** Icelandic sheep, 6 months old, were infected intratracheally with 1.0 ml of 10<sup>4</sup> 50% tissue culture infective doses. Virus was injected into the trachea with a needle (23 gauge) under sedation with xylazine (Xylapan; 0.2 ml intravenously) (34).

**Virus isolation from blood and organs.** Virus was isolated from the buffy coat as described previously (33). The following organs were tested for presence of infectious virus by coculture with SCP cells (24): choroid plexus, cerebrum,

cerebellum, spleen, cervical, mediastinal and mesenteric lymph nodes, bone marrow, and lungs (one sample from each lobe). If no cytopathic effects were observed, the growth medium was passaged twice, with the last observation after 6 weeks of culture.

## RESULTS

**The growth attenuation of the KS1 molecular clone is jointly determined by Vif and CA.** To begin to map the determinants of the attenuated replication phenotype of the KS1 clone in macrophages, a series of recombinants were constructed, using restriction fragments from various parts of KS1 and with KV1772 as a backbone. All recombinant viruses were tested for replication in sheep blood-derived macrophages to see if the impaired replication of KS1 was transferred with the respective fragments. The macrophages were infected with equal amounts of virus (as determined by measuring RT activity), and replication was monitored by taking samples daily and quantifying virus by measuring RT activity or using TaqMan-based real-time PCR.

We first tested a fragment comprising *env*, *tat*, *vif*, and a part of *pol* (VB1) (Fig. 1). There was no effect on replication of the virus in macrophages. Adding further regions of *pol*, *env*, and the long terminal repeat did not have any effect either (clones VB2, VB3, VB5, and VB3 to VB5) (Fig. 1). Only when the two mutations in *gag* were added was the attenuated replication in macrophages transferred (VB4) (Fig. 1). However, replacing a shorter fragment containing the two *gag* mutations did not result in significantly attenuated replication in macrophages (VB6) (Fig. 1). Only when three mutations in *vif* were cloned together with the mutations in *gag* did the virus acquire the attenuated phenotype (VB7) (Fig. 1). The mutations in *vif* did not have an attenuated effect on their own (VB8) (Fig. 1). It thus appears that there is an interaction between the mutations in *gag* and *vif*.

To ask which mutation(s) in *gag* associated with those in *vif*, a series of point mutations were introduced by site-directed mutation of KV1772. One of the two mutations in *gag* was in MA and the other was in CA. To distinguish between these two mutations, they were each constructed separately together with the mutations in *vif*. The virus with the mutation in MA replicated well in macrophages, whereas the virus with the mutation in CA replicated with kinetics similar to those of the KS1 strain, indicating that only the CA mutation was relevant. We then constructed the P-S mutation in Vif separately [strain Vif(P205S)], the mutation in CA separately [strain CA(L120R)], and the two mutations together [strain CA(L120R)-Vif(P205S)]. The replication kinetics of these strains were compared with the two parent strains in choroid plexus cells and macrophages. All strains replicated similarly in choroid plexus cells, whereas in macrophages CA(L120R) was somewhat affected and KS1 and CA(L120R)-Vif(P205S) replicated more slowly and to a lower titer (Fig. 2). Adaptation of KS1 and CA(L120R)-Vif(P205S) to growth in macrophages was observed upon extended passage, but direct reversion mutants were not detected. We are currently searching for possible second-site suppressor mutations.

Interestingly, the P-S mutation in Vif is in a proline-rich region at the C-terminal end of the protein that has been shown to be important for association with cellular membranes and interaction with the Gag precursor Pr55<sup>gag</sup> (3, 9, 10),

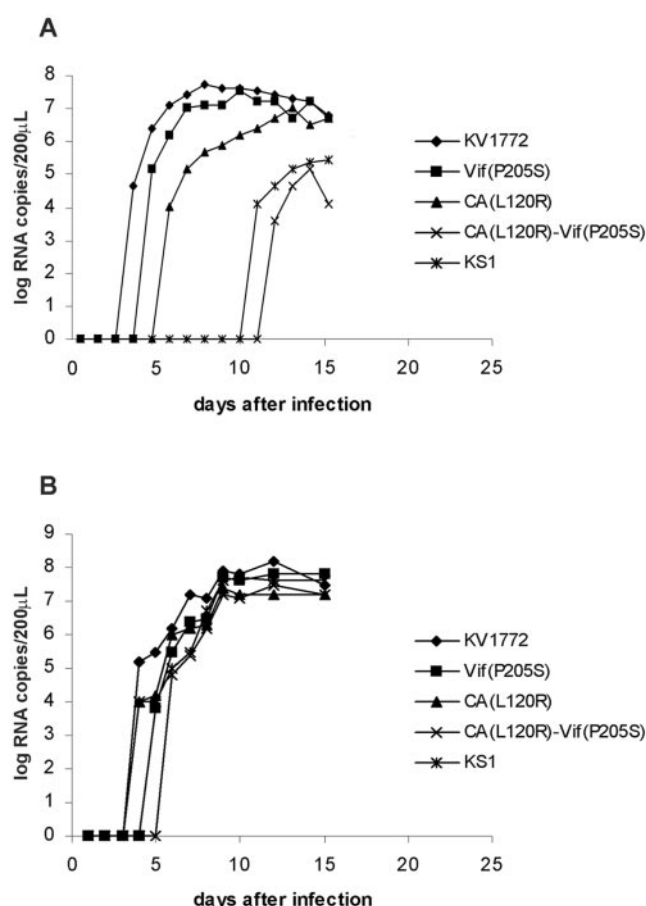


FIG. 2. Replication kinetics of virus production in blood-derived macrophages (A) and sheep choroid plexus cells (B), as measured by TaqMan-based real-time PCR.

whereas the mutation in CA is an L-R mutation in a position corresponding to amino acid 131 in helix 7 in HIV-1 (8, 36).

**In vivo inoculation with the two parental strains and the recombinant virus CA(L120R)-Vif(P205S).** The KS1 strain has repeatedly been shown to be nonpathogenic in sheep, whereas the KV1772 strain is highly pathogenic (33). Of 219 attempts to isolate virus from blood of sheep infected with KS1 in various experiments, virus was isolated only on one occasion. Sequencing of the CA and *vif* genes of this isolate revealed a reversion of the L-R mutation in CA back to L. This isolate grew well in macrophages (Fig. 3).

Three sheep, each 6 months old, were inoculated intratracheally with each of the three virus strains KV1772, KS1, and CA(L120R)-Vif(P205S). Blood was collected for virus isolations once a week for the first 8 weeks and biweekly thereafter until sacrifice after 18 weeks. At sacrifice, virus isolations were attempted from several organs, including brain, lungs, spleen, lymph nodes, and bone marrow. As shown in Table 1, virus was readily recovered from blood and organs of the three sheep that had been inoculated with strain KV1772, whereas no virus was isolated from sheep inoculated with KS1 virus. However, one of the three sheep inoculated with the recombinant CA(L120R)-Vif(P205S) virus was clearly infected. This result was not expected, since the CA(L120R)-Vif(P205S) strain has

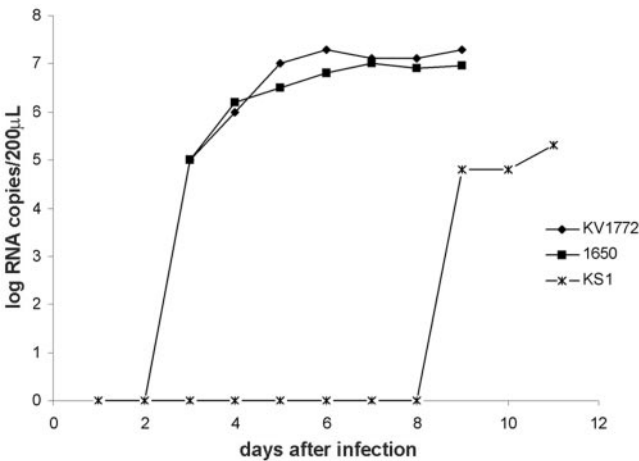


FIG. 3. Replication kinetics of CA revertant (1650), KS1, and 1772 clones in blood-derived macrophages as measured by TaqMan-based real-time PCR.

replication kinetics in macrophages similar to those of the KS1 strain. Sequencing of virus isolates from the infected sheep did not show a reversion of the mutations in CA and *vif*. One of the blood isolates was tested for replication in macrophages, and it showed attenuated replication similar to that of the CA(L120R)-Vif(P205S) strain (data not shown). We then speculated that this particular sheep might not express a putative virus inhibitor. We therefore infected four additional sheep with the virus CA(L120R)-Vif(P205S) and isolated macrophages from each sheep before infection to test replication of the CA(L120R)-Vif(P205S) strain compared to KV1772. Replication of the CA(L120R)-Vif(P205S) strain was attenuated in macrophages from all of the four sheep (data not shown). The sheep were kept for 18 weeks as before, and

attempts were made to isolate virus from the same organs. No virus was isolated from any of the organs, but one isolate was recovered from blood (Table 1).

DISCUSSION

In this study, we traced the attenuated replication and infectivity of the molecular clone KS1 to two simultaneous mutations, one in the CA domain of *gag* and the other in *vif*. The CA mutation affected replication in macrophages somewhat, but with the addition of the *vif* mutation, replication in macrophages was markedly reduced. The mutation in *vif* on its own did not affect replication in macrophages. The finding that a reversion of the mutation in CA in the KS1 clone caused a reversion in phenotype further supports the importance of the mutation in the capsid. The CA mutation is an L-R mutation in amino acid 120 of MVV CA. This is a position corresponding to K131 in helix 7 in the N-terminal domain of HIV-1 CA protein. There is a charged amino acid in this position in all lentiviruses except caprine arthritis encephalitis virus and MVV (36). Charge change mutants of adjacent amino acids in helix 7 in HIV-1 have resulted in alteration in cell tropism that seems to be independent of Ref1/TRIM5α (16). These mutations may therefore define an interaction of the capsid with uncharacterized host proteins.

The mutation in the *vif* gene is a P-to-S mutation in the C terminus, where membrane association, Gag interaction, and Vif multimerization have been mapped for HIV-1; interaction with Gag has also been mapped for caprine arthritis encephalitis virus (3, 9, 10, 25, 37). The interaction with Gag has been mapped to the nucleocapsid part of Gag in vitro, but no interaction has been detected with CA (3, 25). However, Vif has been shown to play a role in the stability of the core of HIV-1 (23), and the interaction between Vif and CA may either be indirect or too transient to be detected by standard methods that are used for detecting protein-protein interactions. It is possible that the two mutations are unrelated and that the effect is additive. However, since the mutation in *vif* did not have a detectible effect on replication on its own, it is tempting to speculate that there is interaction of CA and Vif in the replication of the virus.

Of seven sheep that were inoculated intratracheally with the virus harboring the two mutations [CA(L120R)-Vif(P205S)], one showed multiorgan infection typical of the wild type virus. However, virus isolates from this sheep retained the two mutations and replication characteristics of the mutant virus [CA(L120R)-Vif(P205S)]. It therefore appears that this sheep lacked host restriction. Unfortunately, the sheep is not available for further examination.

It is becoming increasingly clear that a variety of antiretroviral activities have evolved in mammals. The mutations we have described in this report may define a novel host restriction factor that targets the capsid and is counteracted by Vif. Another possibility would be that the CA mutation slows the growth of the virus by destabilizing the core, thus making it more susceptible to host proteins in general. This might exacerbate the effect of APOBEC3 when Vif is attenuated. We are currently addressing the possibility that APOBEC3 knockdown in sheep macrophages rescues growth of the KS1 MVV.

TABLE 1. Frequency of virus isolation from KV1772-, KS1-, and CA(L120R)-Vif(P205S)-infected sheep				
Virus strain and sheep no.	No. positive/no. tested in:			Central nervous system
	Blood	Lymph <sup>a</sup>	Lungs	
KV1772				
2080	10/11	5/5	2/2	1/3
2081	8/11	5/5	2/2	0/3
2082	5/11	2/5	0/2	0/3
KS1				
2074	0/11	0/5	0/2	0/3
2075	0/11	0/5	0/2	0/3
2076	0/11	0/5	0/2	0/3
CA(L120R)-Vif(P205S)				
2077	0/11	0/5	0/2	0/3
2078	2/11	4/5	2/2	0/3
2079	0/11	0/5	0/2	0/3
2089	0/11	0/5	0/2	0/3
2090	0/11	0/5	0/2	0/3
2091	1/11	0/5	0/2	0/3
2092	0/11	0/5	0/2	0/3

<sup>a</sup> Spleen, bone marrow, cervical, mediastinal, and mesenteric lymph nodes.

## ACKNOWLEDGMENTS

This study was supported by the Icelandic Research Council, the University of Iceland Research Fund, and the Icelandic Research Fund for Graduate Students.

We are indebted to Katherine Staskus and Ashley Haase for the KS1 clone and to Svava Högnadóttir, Steinunn Árnadóttir, and Sigurdur Helgason for expert technical help. We thank Reuben Harris for critically reading the manuscript.

## REFERENCES

- Andresdottir, V., X. Tang, G. Agnarsdottir, O. S. Andreson, G. Georgsson, R. Skraban, S. Torsteinsdottir, B. Rafnar, E. Benediktsdottir, S. Matthiasdottir, S. Arnadóttir, S. Hognadóttir, P. A. Pálsson, and G. Petursson. 1998. Biological and genetic differences between lung- and brain-derived isolates of maedi-visna virus. *Virus Genes* **16**:281–293.
- Andresson, O. S., J. E. Elser, G. J. Tobin, J. D. Greenwood, M. A. Gonda, G. Georgsson, V. Andresdottir, E. Benediktsdottir, H. M. Carlsdottir, and E. O. Mantyla. 1993. Nucleotide sequence and biological properties of a pathogenic proviral molecular clone of neurovirulent visna virus. *Virology* **193**:89–105.
- Bouyac, M., M. Courcoul, G. Bertoia, Y. Baudat, D. Gabuzda, D. Blanc, N. Chazal, P. Boulanger, J. Sire, R. Vigne, and B. Spire. 1997. Human immunodeficiency virus type 1 Vif protein binds to the Pr55Gag precursor. *J. Virol.* **71**:9358–9365.
- DesGroseillers, L., and P. Jolicoeur. 1983. Physical mapping of the Fv-1 tropism host range determinant of BALB/c murine leukemia viruses. *J. Virol.* **48**:685–696.
- Desrosiers, R. C., J. D. Lifson, J. S. Gibbs, S. C. Czajak, A. Y. Howe, L. O. Arthur, and R. P. Johnson. 1998. Identification of highly attenuated mutants of simian immunodeficiency virus. *J. Virol.* **72**:1431–1437.
- Gendelman, H. E., O. Narayan, S. Kennedy-Stoskopf, P. G. Kennedy, Z. Ghotbi, J. E. Clements, J. Stanley, and G. Pezeshkpour. 1986. Tropism of sheep lentiviruses for monocytes: susceptibility to infection and virus gene expression increase during maturation of monocytes to macrophages. *J. Virol.* **58**:67–74.
- Georgsson, G., N. Nathanson, P. A. Pálsson, and G. Petursson. 1976. The pathology of visna and maedi in sheep. *Front. Biol.* **44**:61–96.
- Gitti, R. K., B. M. Lee, J. Walker, M. F. Summers, S. Yoo, and W. I. Sundquist. 1996. Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science* **273**:231–235.
- Goncalves, J., P. Jallepalli, and D. H. Gabuzda. 1994. Subcellular localization of the Vif protein of human immunodeficiency virus type 1. *J. Virol.* **68**:704–712.
- Goncalves, J., B. Shi, X. Yang, and D. Gabuzda. 1995. Biological activity of human immunodeficiency virus type 1 Vif requires membrane targeting by C-terminal basic domains. *J. Virol.* **69**:7196–7204.
- Gorrell, M. D., M. R. Brandon, D. Sheffer, R. J. Adams, and O. Narayan. 1992. Ovine lentivirus is macrophagetropic and does not replicate productively in T lymphocytes. *J. Virol.* **66**:2679–2688.
- Harmache, A., M. Bouyac, G. Audoly, C. Hieblot, P. Peveri, R. Vigne, and M. Suzan. 1995. The *vif* gene is essential for efficient replication of caprine arthritis encephalitis virus in goat synovial membrane cells and affects the late steps of the virus replication cycle. *J. Virol.* **69**:3247–3257.
- Harmache, A., P. Russo, F. Guiguen, C. Vitu, M. Vignoni, M. Bouyac, C. Hieblot, M. Pepin, R. Vigne, and M. Suzan. 1996. Requirement of caprine arthritis encephalitis virus *vif* gene for in vivo replication. *Virology* **224**:246–255.
- Harris, R. S., K. N. Bishop, A. M. Sheehy, H. M. Craig, S. K. Petersen-Mahrt, I. N. Watt, M. S. Neuberger, and M. H. Malim. 2003. DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**:803–809.
- Hatzioannou, T., S. Cowan, S. P. Goff, P. D. Bieniasz, and G. J. Towers. 2003. Restriction of multiple divergent retroviruses by Lvl and Ref1. *EMBO J.* **22**:385–394.
- Hatzioannou, T., S. Cowan, U. K. Von Schwedler, W. I. Sundquist, and P. D. Bieniasz. 2004. Species-specific tropism determinants in the human immunodeficiency virus type 1 capsid. *J. Virol.* **78**:6005–6012.
- Hatzioannou, T., D. Perez-Caballero, A. Yang, S. Cowan, and P. D. Bieniasz. 2004. Retrovirus resistance factors Ref1 and Lvl are species-specific variants of TRIM5α. *Proc. Natl. Acad. Sci. USA* **101**:10774–10779.
- Kristbjornsdottir, H. B., V. Andresdottir, V. Svansson, S. Torsteinsdottir, S. Matthiasdottir, and O. S. Andreson. 2004. The *vif* gene of maedi-visna virus is essential for infectivity in vivo and in vitro. *Virology* **318**:350–359.
- Lecossier, D., F. Bouchonnet, F. Clavel, and A. J. Hance. 2003. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* **300**:1112.
- Lockridge, K. M., S. Himathongkham, E. T. Sawai, M. Chienand, and E. E. Sparger. 1999. The feline immunodeficiency virus *vif* gene is required for productive infection of feline peripheral blood mononuclear cells and monocyte-derived macrophages. *Virology* **261**:25–30.
- Mangeat, B., P. Turelli, G. Caron, M. Friedli, L. Perrin, and D. Trono. 2003. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**:99–103.
- Narayan, O. 1983. Role of macrophages in the immunopathogenesis of visna-maedi of sheep. *Prog. Brain Res.* **59**:233–235.
- Ohagen, A., and D. Gabuzda. 2000. Role of Vif in stability of the human immunodeficiency virus type 1 core. *J. Virol.* **74**:11055–11066.
- Petursson, G., N. Nathanson, G. Georgsson, H. Panitch, and P. A. Pálsson. 1976. Pathogenesis of visna. I. Sequential virologic, serologic, and pathologic studies. *Lab. Invest.* **35**:402–412.
- Seroude, V., G. Audoly, P. Gluschkof, and M. Suzan. 2001. Tryptophan 95, an amino acid residue of the Caprine arthritis encephalitis virus Vif protein which is essential for virus replication. *Virology* **280**:232–242.
- Sigurdsson, B., H. Grimsson, and P. A. Pálsson. 1952. Maedi, a chronic progressive infection of sheep's lungs. *J. Infect. Dis.* **90**:233–241.
- Sigurdsson, B., and P. A. Pálsson. 1958. Visna of sheep. A slow demyelinating infection. *Br. J. Exp. Pathol.* **39**:519–528.
- Sigurdsson, B., H. Thormar, and P. A. Pálsson. 1960. Cultivation of visna virus in tissue culture. *Arch. Gesamte Virusforsch.* **10**:368–380.
- Skraban, R., S. Matthiasdottir, S. Torsteinsdottir, G. Agnarsdottir, B. Gudmundsson, G. Georgsson, R. H. Melen, O. S. Andreson, K. A. Staskus, H. Thormar, and V. Andresdottir. 1999. Naturally occurring mutations within 39 amino acids in the envelope glycoprotein of maedi-visna virus alter the neutralization phenotype. *J. Virol.* **73**:8064–8072.
- Staskus, K. A., E. F. Retzel, E. D. Lewis, J. L. Silsby, S. St. Cyr, J. M. Rank, S. W. Wietgreffe, A. T. Haase, R. Cook, and D. Fast. 1991. Isolation of replication-competent molecular clones of visna virus. *Virology* **181**:228–240.
- Strebel, K., D. Daugherty, K. Clouse, D. Cohen, T. Folks, and M. A. Martin. 1987. The HIV 'A' (sor) gene product is essential for virus infectivity. *Nature* **328**:728–730.
- Stremlau, M., C. M. Owens, M. J. Perron, M. Kiessling, P. Autissier, and J. Sodroski. 2004. The cytoplasmic body component TRIM5α restricts HIV-1 infection in Old World monkeys. *Nature* **427**:848–853.
- Torsteinsdottir, S., G. Agnarsdottir, S. Matthiasdottir, B. Rafnar, V. Andresdottir, O. S. Andreson, K. Staskus, G. Petursson, P. A. Pálsson, and G. Georgsson. 1997. In vivo and in vitro infection with two different molecular clones of visna virus. *Virology* **229**:370–380.
- Torsteinsdottir, S., S. Matthiasdottir, N. Vidarsdottir, V. Svansson, and G. Petursson. 2003. Intratracheal inoculation as an efficient route of experimental infection with maedi-visna virus. *Res. Vet. Sci.* **75**:245–247.
- Towers, G., M. Bock, S. Martin, Y. Takeuchi, J. P. Stoye, and O. Danos. 2000. A conserved mechanism of retrovirus restriction in mammals. *Proc. Natl. Acad. Sci. USA* **97**:12295–12299.
- von Schwedler, U. K., T. L. Stemmler, V. Y. Klishko, S. Li, K. H. Albertine, D. R. Davis, and W. I. Sundquist. 1998. Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. *EMBO J.* **17**:1555–1568.
- Yang, S., Y. Sun, and H. Zhang. 2001. The multimerization of human immunodeficiency virus type-1 (HIV-1) Vif protein: a requirement for Vif function in the viral life-cycle. *J. Biol. Chem.* **276**:4889–4893.
- Zhang, H., B. Yang, R. J. Pomerantz, C. Zhang, S. C. Arunachalam, and L. Gao. 2003. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **424**:94–98.

## Paper II



# Evolutionarily conserved and non-conserved retrovirus restriction activities of artiodactyl APOBEC3F proteins

Stefán R. Jónsson<sup>1,2,3,4</sup>, Guylaine Haché<sup>1,2,3</sup>, Mark D. Stenglein<sup>1,2,3</sup>,  
Scott C. Fahrenkrug<sup>3,5</sup>, Valgerdur Andrésdóttir<sup>4</sup> and Reuben S. Harris<sup>1,2,3,\*</sup>

<sup>1</sup>Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, <sup>2</sup>Institute for Molecular Virology, <sup>3</sup>Arnold and Mabel Beckman Center for Transposon Research, Minneapolis, MN 55455, USA, <sup>4</sup>University of Iceland, Institute for Experimental Pathology, Keldur v/Vesturlandsveg, 112 Reykjavík, Iceland and <sup>5</sup>Department of Animal Sciences, University of Minnesota, St Paul, MN 55108, USA

Received August 4, 2006; Revised September 16, 2006; Accepted September 18, 2006

## ABSTRACT

The APOBEC3 proteins are unique to mammals. Many inhibit retrovirus infection through a cDNA cytosine deamination mechanism. HIV-1 neutralizes this host defense through Vif, which triggers APOBEC3 ubiquitination and degradation. Here, we report an APOBEC3F-like, double deaminase domain protein from three artiodactyls: cattle, pigs and sheep. Like their human counterparts, APOBEC3F and APOBEC3G, the artiodactyl APOBEC3F proteins are DNA cytosine deaminases that locate predominantly to the cytosol and can inhibit the replication of HIV-1 and MLV. Retrovirus restriction is attributable to deaminase-dependent and -independent mechanisms, as deaminase-defective mutants retain significant anti-retroviral activity. However, unlike human APOBEC3F and APOBEC3G, the artiodactyl APOBEC3F proteins have an active N-terminal DNA cytosine deaminase domain, which elicits a broader dinucleotide deamination preference, and they are resistant to HIV-1 Vif. These data indicate that DNA cytosine deamination; sub-cellular localization and retrovirus restriction activities are conserved in mammals, whereas active site location, local mutational preferences and Vif susceptibility are not. Together, these studies indicate that some properties of the mammal-specific, APOBEC3-dependent retroelement restriction system are necessary and conserved, but others are simultaneously modular and highly adaptable.

## INTRODUCTION

Expression of many of the human APOBEC3 (A3) proteins has been shown to inhibit the infective potential and mobility of a broad and growing number of retroviruses and retrotransposons [reviewed by (1–4)]. Humans encode seven A3 proteins, *Homo sapiens* (Hs) HsA3A, HsA3B, HsA3C, HsA3DE, HsA3F, HsA3G and HsA3H, in tandem on chromosome 22 (2,5–7). HsA3A has recently been shown to inhibit the mobility of both long terminal repeat (LTR) and non-LTR retrotransposons (8–10). HsA3B can also inhibit L1 and Alu retrotransposition, as well as the replication of SIV and to a lesser extent HIV-1 (8–13). HsA3C potently inhibits SIV, but it has shown little activity against other substrates (11,13). HsA3DE was recently shown to possess weak antiviral activity (14). HsA3G was the first member of this family to be associated with HIV restriction (15). HsA3F and HsA3G are both capable of potently inhibiting a variety of exogenous and endogenous retroelements (9,10,12,13,15–25). Finally, although HsA3H elicited DNA cytosine deaminase activity, it was unable to restrict SIV or HIV-1 replication (26). Several simian (e.g. chimpanzee), one carnivore (cat) and one rodent (mouse) APOBEC3 protein have also been shown to possess retroelement restriction activities [e.g. (20,25,27–30)]. APOBEC3 proteins from other mammals have yet to be examined.

Three themes appear to be emerging from these studies. First, the A3 proteins deaminate cytosines to uracils (C→U) within single-strand DNA (ssDNA). This property enables the A3 proteins to target the cDNA replication intermediates of all of the aforementioned retroviruses and retrotransposons. Second, retroelement restriction is mediated by at least two distinct mechanisms—by retroviral cDNA cytosine deamination (the hallmark activity of this family of proteins) and by a

\*To whom correspondence should be addressed. Tel: +1 612 624 0457; Fax: +1 612 625 2163; Email: rsh@umn.edu

deamination-independent mechanism that is not well understood [e.g. (12,31–34)]. Finally, many of the retroelements (especially the retroviruses) that have been examined in detail can evade A3-dependent restriction. For instance, the virion infectivity factor (Vif) of HIV-1 and SIV recruits a cellular ubiquitin ligation complex to purge cells of A3G, the Bet protein of several different foamy viruses appears to directly bind and neutralize A3G and other retroviruses have simply evolved to exclude A3 proteins from nascent virions (e.g. MLV and HTLV) (13,28–30,33,35–39).

Strong evidence indicates that the conflict between host A3 proteins and invasive retroelements is ancient. Phylogenetic analyses based on representative vertebrate genome sequences indicate that the A3 proteins are at least as old as the mammalian lineage, because rodents encode one and primates seven A3 proteins (2,6). Other vertebrates, such as birds and fish do not have A3 proteins *per se*, but they do encode activation-induced deaminase (AID), an A3 orthologue that uses DNA cytosine deamination to trigger immunoglobulin gene hypermutation and isotype switch recombination [recently reviewed by (40–42)]. Comparative studies of A3 proteins from humans and non-human primates, New World monkeys, such as the tamarin and the woolly monkey, have demonstrated that the mammalian A3 proteins have been under a strong positive selection for at least 33 million years (26,43,44). The strong and likely ongoing positive selection and the unparalleled A3 gene expansion from one in rodents to seven in primates combine to suggest that the A3 proteins form a highly flexible and adaptable innate host defense system, which may very well be capable of readily adapting to new and potentially invasive retroelements.

The evolutionary gap between rodents and primates is ~90–100 million years [e.g. (45)]. This large genetic distance enables some comparative studies, but it limits others. Therefore, to close some of this distance, to enable more extensive comparative studies, and to examine potentially novel A3-virus conflicts, we have cloned and characterized A3 proteins from representative artiodactyls: cattle (*Bos taurus*; Bt), sheep (*Ovis aries*; Oa) and pigs (*Sus scrofa*; Ss). These studies focused on artiodactyl A3 proteins that were predicted to be similar in size (ca. 400 amino acids) and domain organization (two conserved zinc-binding deaminase domains) to the mouse *Mus musculus* (Mm) A3 protein and the well-characterized HsA3F and HsA3G proteins. The results of our studies emphasize the importance of DNA cytosine deamination in retrovirus restriction, but they also highlight the existence of a conserved deaminase-independent restriction mechanism. Moreover, non-conserved properties, such as Vif susceptibility, active site location and local mutational preferences combine to suggest mechanistic flexibilities that the APOBEC3 proteins might employ while adapting to new and potentially threatening genetic challenges.

## MATERIALS AND METHODS

### Artiodactyl A3F cDNA sequences and expression plasmids

*Escherichia coli* expression plasmids were derived from pTrc99A (GE Healthcare). The SsA3F expression

construct was made by PCR amplifying a cDNA [GB# BI346898 (46)] using 5'-NNNNGAGCTCAGG-TACCACCATGGATCCTCAGCGCCTGAGAC and 5'-NNNGTCTGACTCATCTCGAGTCACTTCTTGATG, digesting the PCR product with KpnI and SalI, and ligating it into a similarly digested pTrc99A. The BtA3F expression construct was made in the same manner by PCR amplifying a cDNA [GB# BE684372 (47)] using oligonucleotides 5'-NNNNGAGCTCAGGTACCACCATGCAACCAGCCTAC-CGAGGC and 5'-NNNNGTCTGACCTAAATTGGGGCCGT-TAGGAT. OaA3F was obtained by first amplifying a fragment by degenerative PCR using oligonucleotides 5'-TWYR-TVTCCTGGAGCCCCTG and 5'-CCRKWWGTWGTAGA-GGCGR, which were made to conserved regions of APOBEC3 proteins using template cDNA from sheep macrophages. The amplified fragment was sequenced and used to make OaA3F-specific oligonucleotides for 3' RACE to obtain the remainder of the OaA3F coding sequence, 5'-AACCAG-GTCTATGCTGGGACT and 5'-CTGGGGATGTACCAGATGTG were used with an oligo dT primer 5'-AAGCAGTG-GTAACAACGCAGAGTACT<sub>30</sub>VN. The OaA3F expression plasmid was made by amplification from cDNA of sheep macrophages using oligonucleotides 5'-NNNNGAGCTCAG-GTACCACCATGCCCTGGATCAGCGACCAC and 5'-NNNGTCTGACCTAAGTCGGCGCCGTCAAGAT in the same manner as the BtA3F and SsA3F constructs. E→Q zinc-binding domain mutants were constructed using standard site-directed mutagenesis protocols (Stratagene). The following oligonucleotides were used: SsA3F (E87→Q) 5'-CCGACCCGCCCTGCCACGCCAGCTCTGCTTCCTCTCTTGT and 5'-ACAAGAGAGGAAGCAGAGCTGGGCGTGCGA-GGGCGGGTCGG; SsA3F (E267→Q) 5'-ACAAGAAAAAGCGACATGCACAAATTCGTTTTATTGACAAG and 5'-CTTGTCATAATAAACGAATTTGTGCATGTGCTTTTTCTTGT; BtA3F (E80→Q) 5'-GTGGGACTCGCTGCCACACCCAACTCC;GCTTCCTGTCTTGG and 5'-CCAAGACAGGAAGCGGAGTTGGGTGTGGCAGCGAGTCCCAC; BtA3F (E260→Q) 5'-ACAAGAAGCAGCGGCATGCACAAATTCGCTTTATTGACAAG and 5'-CTTGTCATAAAGCGAATTTGTGCATGCCGCTGCTTCTTGT; OaA3F (E67→Q) 5'-CTGGGACTCACTGCCACAGCCAAACGCCGCTTCCTGTCTTGG and 5'-CCAAGACAGGAAGCGGCGTTGGCTGTGGCAGTGAGTCCCAG; OaA3F (E247→Q) 5'-ACAAGAAGCAGCGGCATGCACAAATTCGCTTTATTGACAAG and 5'-CTTGTCATAAAGCGAATTTGTGCATGCCGCTGCTTCTTGT. Eukaryotic expression plasmids were derived from pcDNA3.1(+) (Invitrogen) by subcloning the KpnI/SalI-flanked A3 cDNAs from the *E.coli* expression plasmids.

### Additional DNA constructs

HsA3F, HsA3G and HsAID expression plasmids were described previously (31,48). MmA3 was made in the same manner as the artiodactyl constructs using oligonucleotides 5'-NNNNGAGCTCGGTACCACCATGGGACCATCTGTCTGGGA and 5'-NNNNGTCTGACATCAAGACATCGGGGGTCCAAGCTG. E→Q zinc-binding domain mutants were constructed using standard site-directed mutagenesis protocols (Stratagene). The following oligonucleotides were

used: HsA3G (E67→Q) 5'-CTTAAGTACCACCCACAGATGAGATTCTTCC and 5'-GGAAGAATCTCATCTGTGGGTGGTACTTAAG; HsA3G(E259→Q) 5'-CCTTGAA-GGCCGCCATGCACAGCTGTGCTTCCTGG and 5'-CCA-GGAAGCACAGCTGTGCATGGCGGCCTTCAAGG; HsA3F(E67→Q) 5'-CTGAGCACCACGCACAAATGTGCTTCCTC and 5'-GAGGAAGCACATTTGTGCGTGGTGCTCAG; HsA3F(E251→Q) 5'-CCCATTGTCATGCACAAAGGTGCTTCCTC and 5'-GAGGAAGCACCTTTGTGCATGACAATGGG; MmA3(E73→Q) 5'-CCAGTATAAAAAGCAGATTTGAGCGTGGATGTTGTCCTTG and 5'-CAAG-GACAACATCCACGCTCAAATCTGCTTTTATACTGG; MmA3(E290→Q) 5'-CTTATCAAGGAAGAGGATTTGTGCATGCTGTTTGCCTTTC and 5'-GAAAGGCAAACAGCATGCACAAATCCTCTTCCTTGATAAG. Eukaryotic expression plasmids were derived from pcDNA3.1(+) (Invitrogen) and pEGFP-N3 (Clontech) by inserting KpnI and SalI or SacI and SalI-flanked A3 cDNAs from the *E.coli* expression plasmids. The coding sequences of HsAID, HsA3F, HsA3G and MmA3 used in this study are identical those represented by GenBank accession nos NP\_065712, NP\_660341, NP\_068594 and AAH03314, respectively. The coding sequence of HsA3B has two amino acid substitutions, W228L and D316N (12), which distinguish it from GenBank NP\_004891.

#### Amino acid alignments and phylogenetic analyses

Protein alignments and phylogenetic analyses of the conserved, zinc-binding deaminase cores were done using the Clustal W software (49). A rooted phylogenetic tree was constructed using the neighbor-joining method. The GenBank accession numbers of the sequences used in these comparisons were AAI01974 (HsA1), NP\_112436 (MmA1), NP\_037039 (RnA1), NP\_006780 (HsA2), XP\_538909 (CfA2), NP\_001029527 (BtA2), NP\_033824 (MmA2), XP\_217334 (RnA2), XP\_528775 (PtAID), NP\_065712 (HsAID), NP\_033775 (MmAID), XP\_575660 (RnAID), NP\_001033771 (BtAID), NP\_663745 (HsA3A), NP\_004891 (HsA3B), NP\_055323 (HsA3C), Q96AK3 (HsA3DE), NP\_660341 (HsA3F), NP\_068594 (HsA3G), NP\_861438 (HsA3H), AAT44392 (PtA3G), ABD72578 (PtA3H), AAH03314 (MmA3) and NP\_001028875 (RnA3).

#### Rif<sup>R</sup> mutation assays

The intrinsic DNA cytosine deaminase activity of the A3 proteins was assayed by quantifying the accumulation of Rif<sup>R</sup> mutants in *ung*-deficient *E.coli* [e.g. (31,48)]. Briefly, dilutions of A3- or vector control-expressing single colonies were grown to saturation in a rich bacterial growth medium containing 100 µg/ml ampicillin and 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Aliquots were spread to plates containing rifampicin (100 µg/ml; Sigma) to select for Rif<sup>R</sup> mutants. Dilutions were spread to plates containing ampicillin to determine the number of viable cells. Mutation frequencies were calculated as the number of Rif<sup>R</sup> mutants per 10<sup>7</sup> viable cells. For each experimental condition, 8 or 12 independent transformants were analyzed.

#### Fluorescence microscopy

Cells were visualized on a Zeiss Axiovert 200 microscope at 400× total magnification. A total of 7500 HeLa cells were seeded on LabTek chambered coverglasses (Nunc). After 24 h of incubation, these cells were transfected with 200 ng of the pEGFP-N3-based DNA constructs. After an additional 24 h of incubation, images of the live cells were collected. HsA3B, HsA3F and HsA3G-eGFP fusion constructs were reported previously (12).

#### Green fluorescent protein (GFP)-based retrovirus infection assays

293T cells were grown in DMEM supplemented with 10% FBS, penicillin and streptomycin. Viruses were produced by lipid-mediated transfection (Fugene; Roche) of 50% confluent 293T cells with the following plasmids. HIV-GFP: 0.22 µg of CS-CG (50), 0.14 µg of pRK5/Pack1(Gag-Pol), 0.07 µg of pRK5/Rev, 0.07 µg of pMDG (VSV-G Env) and 0.5 µg of an A3 expression plasmid or an empty vector control, as described previously (31,51). In some experiments, an HIV-1 Vif or a ΔVif control plasmid was included (0.5 µg). MLV-GFP was produced similarly, except the proviral plasmid was pM3P-GFP and a MLV gag/pol construct was used (18,51). After 48 h of incubation, the virus-containing supernatants were clarified by low speed centrifugation, filtered (0.45 µm), and quantified using a reverse transcriptase activity-based enzyme-linked immunosorbent assay (ELISA) (Cavidi Tech). Reverse transcriptase-normalized supernatants were applied to fresh 293T cells, and infection was allowed to proceed for 72 h. Infectivity (GFP fluorescence) was measured by flow cytometry (FACSCalibur, BD Biosciences). For experiments requiring the recovery of retroviral DNA for hypermutation analyses, the viral supernatants were treated with 50 U/ml DNase (Sigma) prior to 293T cell infection.

#### Retroviral DNA sequence analyses

Genomic DNA was incubated with DpnI to remove possible contaminating input CS-CG plasmid DNA. GFP was amplified using high-fidelity Phusion polymerase (Finnzymes) and CS-CG-specific primers, 5'-CGTGTACGGTGGGAGGTCTA and 5'-TTGGTAGCTGCTGTGTTGCT. PCR products were cloned and sequenced as described previously (51). Mutational analyses were performed using Sequencher software (Gene Codes Corp.).

#### GenBank accession numbers

The OaA3F cDNA sequence has been assigned GenBank DQ974645. The GenBank EST entries for BtA3F (BE684372) and SsA3F (BI346898) have been updated to include the full cDNA sequence (DQ974646 and DQ974647, respectively).

## RESULTS

#### The double deaminase domain APOBEC3F proteins of artiodactyls

All known A3 proteins have either one or two conserved, zinc-binding deaminase domains, consisting of amino acids

HXE-X<sub>23-28</sub>-PCX<sub>2-4</sub>C (X can be any amino acid) (2,5-7). The histidine and the two cysteines coordinate zinc and the glutamate participates directly in the C→U deamination reaction. NCBI BLAST searches using the human and mouse A3 deaminase domains as query polypeptides revealed several artiodactyl ESTs, which suggested the presence of at least one A3 protein in cattle and pigs. Corresponding cDNA clones were obtained, sequenced and shown to encode A3 proteins with two putative zinc-binding, cytosine deaminase domains (Figure 1A and Supplementary Figures S1 and S2; Materials and Methods). The orthologous sheep double domain A3 cDNA sequence was obtained using a combination of degenerate PCR and RACE (Figure 1A and Supplementary Figures S1 and S2). All three of these A3 proteins were similar in size to the 373 amino acid HsA3F protein, except the pig A3 protein, which was slightly longer due to a unique C-terminal, serine-rich extension (Figure 1A and Supplementary Figures S1 and S2).

Nomenclature standards dictate that protein names should be assigned based on the closest human orthologue [e.g. (52,53)]. However, amino acid comparisons showed that the N- and the C-terminal artiodactyl deaminase domains were most similar to different human A3 proteins (data not shown). To resolve this ambiguity and to facilitate name assignments, we named the artiodactyl double domain proteins after 'the human double domain A3 protein with the highest degree of active site identity' (see below for experimental demonstrations). The active sites of these artiodactyl A3 proteins were 56-62% identical to HsA3F (Figure 1B and Supplementary Figure S1). With the exception of HsA3E, which has equivalent identity (but no demonstrated activity), all of the other human A3 proteins had less identity. Thus, we have named the double domain deaminase proteins from cow, sheep and pig after HsA3F and, hereafter, will refer to them as BtA3F, OaA3F and SsA3F, respectively. Amino acid alignments of the active deaminase domains plus five residues on each side showed that the cow and sheep A3F active sites are 78% identical (Figure 1B and Supplementary Figures S1 and S2). Both the cow and the sheep proteins shared a lower level of identity with the pig protein (56%). These identity differences are consistent with the estimated times of divergence from their last common ancestor, as cows and sheep were estimated to have diverged from each other 14-20 million years ago, whereas pigs diverged from this lineage approximately 35-55 million years ago [e.g. (54)].

### Phylogenetic analyses of the artiodactyl A3F proteins

As described above, all A3 proteins have either one or two conserved, zinc-binding (Z) deaminase domains [e.g. Figure 1A]. These domains cluster into three distinct phylogenetic groups: Z1a, Z1b or Z2 (7). The human double Z domain proteins, HsA3F and HsA3G, have a Z1a/Z1a and a Z1a/Z1b organization, respectively, whereas the MmA3 protein has a Z1a/Z2 organization. Interestingly, all three of the artiodactyl A3F proteins have a Z1a/Z2 organization (Figure 1C).

The active sites of most of the human DNA cytosine deaminase-competent A3 proteins can be classed as Z1 [e.g. HsA3F is Z1a and HsA3G is Z1b (31,32)]. Recently,

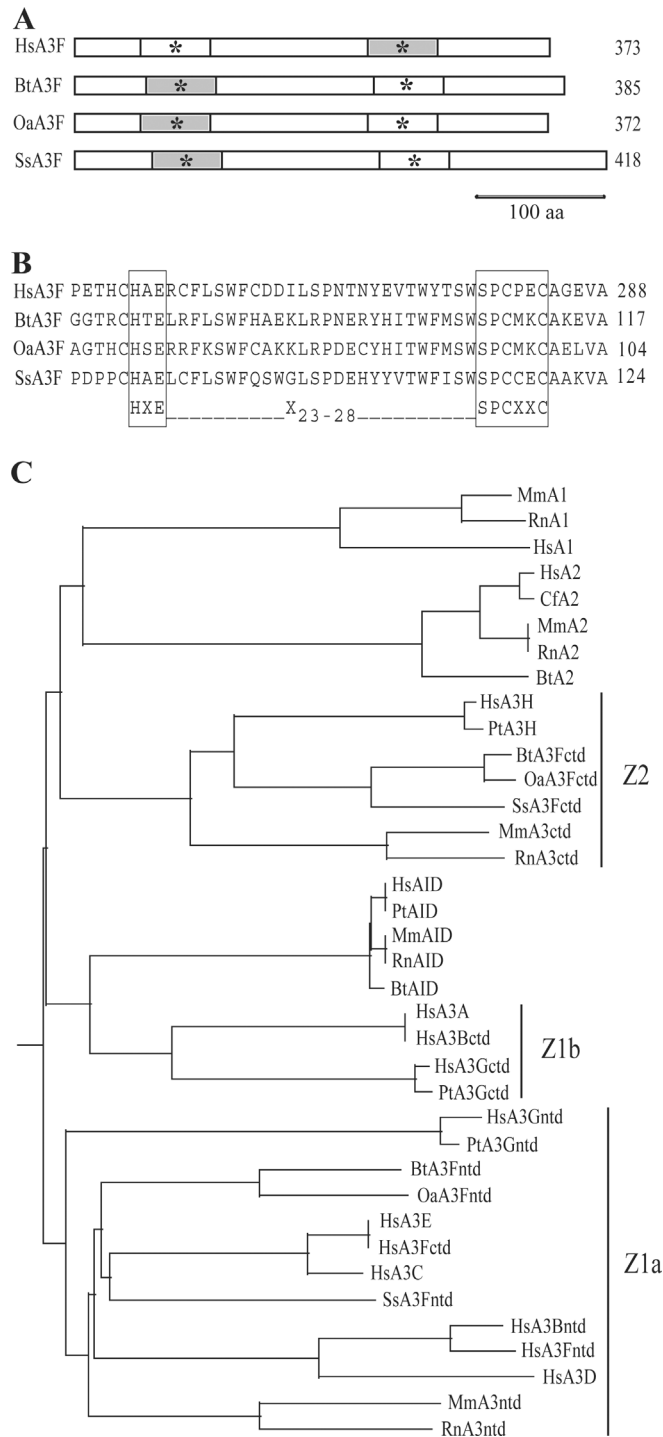
the only human protein with a Z2 designation, HsA3H, was also shown to possess DNA cytosine deaminase activity (26). It is not clear, which Z domain(s) of MmA3 is active (addressed below). Thus, the N-terminal Z1a or the C-terminal Z2 domain of the artiodactyl A3F proteins appeared to have the potential to catalyze DNA cytosine deamination.

### The artiodactyl A3F proteins catalyze DNA cytosine deamination

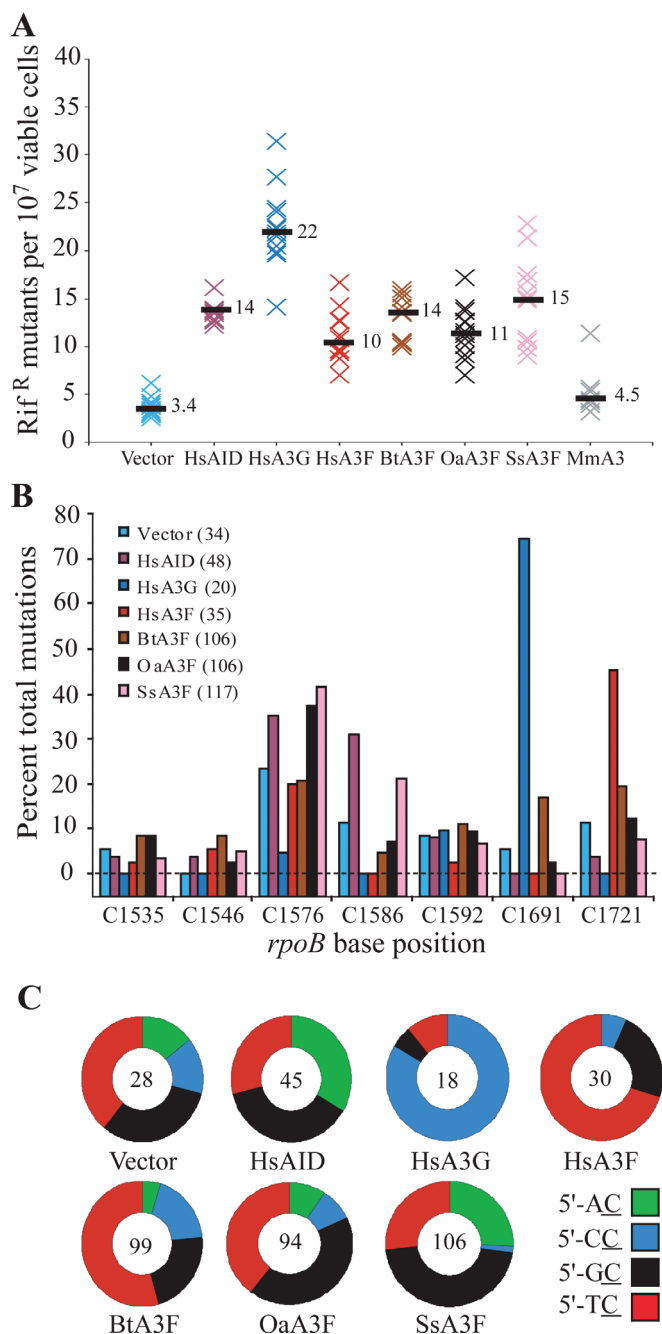
To test whether the artiodactyl A3F proteins have the capacity to deaminate cytosines within ssDNA, the intrinsic mutator activity of these proteins was monitored using an *E.coli*-based mutation assay. Rif<sup>R</sup> is attributable to base substitution mutations in the *E.coli* RNA polymerase B (*rpoB*) gene, and it occurs in approximately 1 of every 5 million bacterial cells. We have found previously that expression of several A3 family members, including HsAID, rat APOBEC1, HsA3C, HsA3F and HsA3G, can accelerate the accumulation of Rif<sup>R</sup> mutations from a few- to several 100-fold (48,51,55). The mutator phenotype is accounted for by a pronounced C/G→T/A transition bias within *rpoB*. This assay therefore provides a robust measure of intrinsic DNA cytosine deaminase activity.

Expression of each of the artiodactyl A3 proteins increased the Rif<sup>R</sup> mutation frequency in *E.coli* from 3- to 7-fold, levels that were higher than those attributable to HsA3F but slightly lower than those caused by HsA3G (Figure 2A). Curiously, expression of MmA3 failed to cause an *E.coli* mutator phenotype, despite the fact that it is clearly active and capable of deaminating the cDNA of a variety of retroelements [e.g. (20,27) and below]. It is not clear why MmA3 is inactive in this system, yet active in others.

Artiodactyl A3F DNA cytosine deamination preferences were examined by sequencing the *rpoB* gene of at least 100 independent Rif<sup>R</sup> mutants (Figure 2B and C). In contrast to HsA3F and HsA3G, which preferentially deaminate cytosines at *rpoB* nucleotide positions 1721 and 1691, 5'-TC and 5'-CC, respectively (48,51), the artiodactyl A3F proteins showed less biased *rpoB* mutation spectra. OaA3F preferentially deaminated cytosine 1576, which is part of a 5'-GC dinucleotide. SsA3F also preferred cytosine 1576. However, SsA3F also clearly deaminated cytosine 1586, which is part of a 5'-AC dinucleotide. Interestingly, these two cytosines, C<sub>1576</sub> and C<sub>1586</sub>, are also preferred by HsAID [Figure 2B and (48,55)]. The fact that HsAID and SsA3F appear to share a 5'-purine-C deamination preference suggests that their common ancestor [perhaps an ancient AID (40)] may have had a similar target preference. However, several groups have reported that even subtle amino acid substitutions in HsA3 proteins can dramatically alter local DNA cytosine deamination preferences, and therefore it is not surprising that most of the present day A3 proteins show non-AID-like mutational preferences (13,31,56). For instance, BtA3F did not appear to have any prominent *rpoB* local mutation preference, as increased levels of C/G→T/A mutation were apparent at several sites. In conclusion, all three of the artiodactyl A3F proteins are capable of deaminating DNA cytosines to uracils, which triggers a corresponding shift in the pattern of C/G→T/A transition mutations within the *rpoB* mutation substrate.



**Figure 1.** A comparison of artiodactyl A3F proteins. (A) A schematic of HsA3F, BtA3F, OaA3F and SsA3F. The conserved, zinc-binding deaminase domains are boxed (\*) and those that are catalytically active are additionally shaded. The numbers on the right indicate the total polypeptide length. (B) HsA3F, BtA3F, OaA3F and SsA3F active site amino acid alignments [shaded regions from (A)]. The conserved HXE and PCXXC motifs are boxed. Amino acid positions are indicated on the right. (C) A neighbor-joining phylogenetic tree indicating the evolutionary relationship of several representative mammalian A3 family members. Branch lengths are proportional to the number of amino acid differences. Comparisons were done using the conserved Z domain amino acids, plus five additional residues on either side. The Z1a, Z1b and Z2 phylogenetic clusters are indicated. HsA3D and HsA3E represent the N- and C-terminal domains of HsA3DE. See the text for additional details. Bt, *B.taurus* (cow); Cf, *Canis familiaris* (dog); Hs, *H.sapiens* (human); Oa, *O.aries* (sheep); Pt, *Pan troglodytes* (chimpanzee); Rn, *Rattus norvegicus* (rat); Ss, *S.scrofa* (pig).



**Figure 2.** DNA cytosine deaminase activity of the artiodactyl A3F proteins in *E. coli*. (A) Rif<sup>R</sup> mutation frequencies for 12 independent *E. coli* cultures expressing a vector control (light blue), HsAID (purple), HsA3G (dark blue), HsA3F (red), BtA3F (brown), OaA3F (black), SsA3F (pink) or MmA3 (gray). Each data point corresponds to the mutation frequency obtained from a single culture, and the median mutation frequency for each condition is shown (horizontal bar). (B) A histogram summarizing the C/G→T/A transition mutations detected in *rpoB*. Only cytosines that had greater than two mutations are shown. Apart from C<sub>1586</sub>, all of the cytosines are located in the non-template strand of the *rpoB* gene. The number of independent Rif<sup>R</sup> mutants that were sequenced is indicated in parentheses in the legend. For purposes of presentation the Y-axis extends below zero (a dotted line marks the actual base-line), and the histogram bars follow the color scheme (A). (C) Pie graphs depicting the frequency that each of the four dinucleotides was targeted by the indicated A3 protein. The total number of independent sequences analyzed is shown in the center of each graph, and the deaminated cytosine is underlined in the legend. The dinucleotide wedges of each pie are colored as indicated.

### The artiodactyl A3F proteins are predominantly cytoplasmic

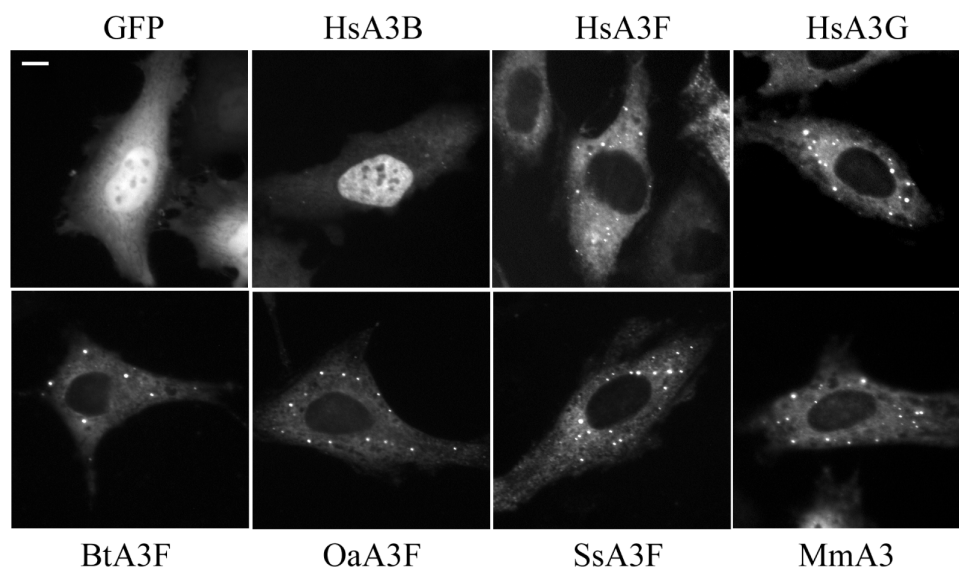
As an initial step toward understanding the potential retroelement targets of the artiodactyl A3F proteins, the sub-cellular distribution of these proteins was determined. A3-GFP constructs were transfected into HeLa cells and the sub-cellular localization of the A3 proteins was determined by live cell fluorescence microscopy. In agreement with prior work, HsA3B and a GFP-only control localized to the nucleus and cell-wide, respectively (12). In contrast, the artiodactyl A3F proteins and MmA3 appeared predominantly cytoplasmic (Figure 3). Many cells contained brightly fluorescing, punctate cytoplasmic aggregations, which may represent P bodies [e.g. (57)]. The significance of the cytoplasmic punctae remains to be determined. Nevertheless, the cytoplasmic localization pattern is nearly identical to that of HsA3F and HsA3G suggesting that the property of localizing to the cytoplasm is conserved and that the artiodactyl A3F proteins might function similarly to inhibit the replication of LTR-dependent retroviruses, such as HIV-1 or MLV [Figure 3; compare with (9,12,16,58)]. One should note, however, that the present data do not exclude the possibility that one or more of these A3 proteins might also possess the nucleo-cytoplasmic shuttling capability of AID, which also appears predominantly cytoplasmic but is clearly capable of entering and exiting the nuclear compartment where it triggers antibody gene diversification processes (59,60).

### Retrovirus restriction by artiodactyl A3F proteins

A clear trend in the genetic conflict between A3 proteins and retroelements is that an A3 from a given host is either neutralized or avoided by retroelements that are specific to the host species. For instance, HIV-1 Vif counteracts both HsA3F and HsA3G, and SIV<sub>agm</sub> Vif inhibits African green monkey A3G [e.g. (51,61–64)]. In many instances, cross-species comparisons enable potential species-specific mechanisms of neutralization to be avoided and the restrictive potential of A3 proteins to be studied. We therefore asked whether the artiodactyl A3F proteins could inhibit the infectivity of HIV-1 and MLV-based retroviruses. In these systems, a *GFP* gene embedded in proviral DNA provides a measure of both transfection efficiency (which correlates directly with virus production levels) and viral infectivity [e.g. (18,51)].

Expression of HsA3F and HsA3G caused 4- and 24-fold reductions in the infectivity of HIV-GFP, in agreement with previous studies [Figure 4A and (51,64)]. MmA3 was also capable of strongly inhibiting HIV-GFP. In comparison, expression of BtA3F, OaA3F or SsA3F caused 30-, 8- and 29-fold decreases in the infectivity of HIV-GFP, respectively (Figure 4A). These potent anti-HIV activities demonstrated that the artiodactyl A3F proteins possess at least one retrovirus restriction activity. These results further imply that the artiodactyl A3F proteins are able to specifically associate with the HIV-1 Gag-genomic RNA complex and thereby gain access to assembling virus particles (addressed further below).

Expression of MmA3 has little effect on the infectivity of MLV, presumably because MLV excludes (or simply avoids) this A3 protein [Figure 4B and (13,27,37,39)]. In contrast, HsA3F and HsA3G inhibit the infectivity of MLV-based



**Figure 3.** Sub-cellular distribution of the artiodactyl A3F proteins in comparison to the orthologous human and mouse A3 proteins. HeLa cells showing localization of the indicated, GFP-tagged A3 proteins or a GFP-only control. The scale bar indicates 10  $\mu$ m.

retroviruses, but to a lesser extent than HIV lacking Vif [Figure 4B and, e.g. (18,51,64)]. Therefore, to ask whether the artiodactyl A3F proteins possess similar restriction potentials, the infectivity of MLV-GFP produced in the presence of these A3 proteins was monitored. Interestingly, similar to HsA3F or HsA3G, expression of the artiodactyl A3F proteins reduced the infectivity of MLV-GFP by 2- to 4-fold. Thus, the HIV-GFP and MLV-GFP infectivity data combined to suggest that the artiodactyl A3F proteins have a relatively broad retrovirus restriction potential.

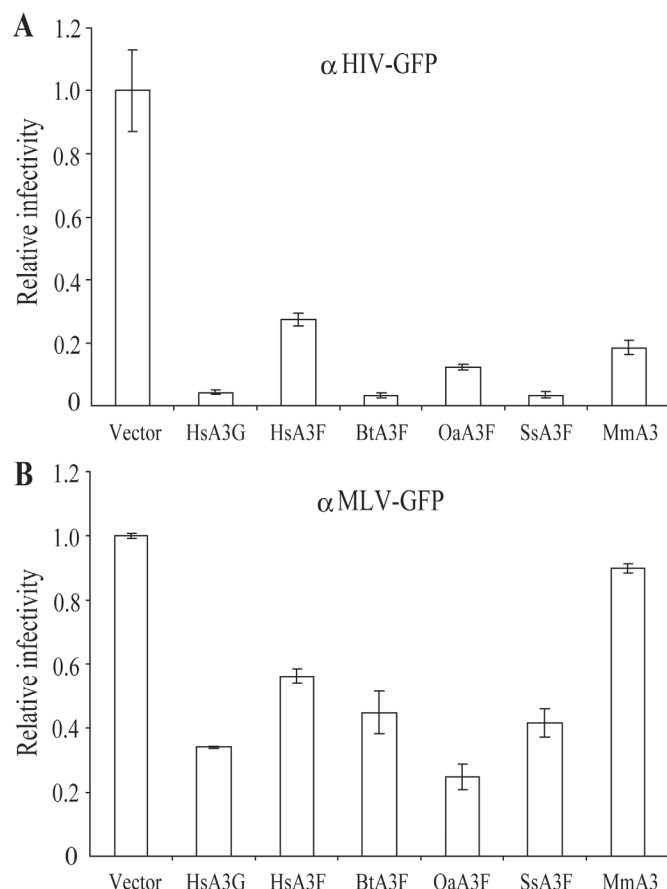
**The N-terminal zinc-binding, deaminase domain of the artiodactyl A3F proteins catalyzes C $\rightarrow$ U deamination, and this activity is necessary for full levels of retrovirus restriction**

All of the double domain deaminases thus far characterized have catalytically competent C-terminal Z domains and apparently inert N-terminal Z domains (10,12,31,32,34). Most experiments have focused on HsA3B, HsA3F and HsA3G, DNA cytosine deaminases that have Z1a- or Z1b-type, C-terminal active sites (Figure 1). However, HsA3H, which has a single Z2-type deaminase motif, was also shown to possess DNA cytosine deaminase activity (26). Thus, it was possible that either the N- or the C-terminal Z domain (or both) of the artiodactyl A3F proteins would be catalytically active.

To begin to work-out the mechanism of retrovirus restriction by artiodactyl A3F proteins and to test whether the N- or the C-terminal (or both) Z domain of these proteins catalyzes DNA cytosine deamination, the conserved glutamate (E) of each active site was changed to glutamine (Q) and the resulting mutants were tested for HIV-GFP restriction activity. As reported previously, the glutamate of both the N- and the C-terminal Z domain of HsA3G contributed to inhibiting HIV-1 infectivity, but the C-terminal catalytic glutamate appeared to be more important [Figure 5A and (34,65,66)]. In contrast, both the N- and the C-terminal

BtA3F Z domain E $\rightarrow$ Q mutants appeared to retain full levels of anti-HIV activity. Interestingly, the N-terminal OaA3F and SsA3F Z domain E $\rightarrow$ Q mutants were less able than the corresponding C-terminal domain mutants to inhibit the infectivity of HIV-GFP, a result particularly clear for SsA3F (Figure 5A). These data were essentially the inverse of the HsA3F and HsA3G E $\rightarrow$ Q mutant studies, and they suggested that the N-terminal, Z domain of these proteins catalyzes retroviral cDNA C $\rightarrow$ U deamination. MmA3 was clearly distinct, as both the N- and the C-terminal Z domain glutamates were required for HIV-GFP restriction (Figure 5A).

Although both the N- and the C-terminal Z domain E $\rightarrow$ Q mutants of the human and the artiodactyl A3 proteins showed significant levels of anti-retroviral activity, we surmised that bona fide catalytic site mutants should be unable to catalyze retroviral cDNA C $\rightarrow$ U deamination [although they may still inhibit retroviral infectivity; e.g. (32)]. Minus strand uracils template the incorporation of plus strand adenines, ultimately manifesting as retroviral plus strand G $\rightarrow$ A hypermutations. Therefore, to directly test which Z domain(s) catalyzes DNA cytosine deamination and to gain additional insight into the artiodactyl A3F retrovirus restriction mechanism, the *GFP* gene from the aforementioned HIV-GFP infectivity experiments was amplified by high-fidelity PCR, cloned and sequenced. HIV-GFP produced in the presence of a control vector showed a low base substitution mutation frequency, 0.00014 mutations per base, which is attributable to errors in RT and PCR (Figure 5B). In contrast, viruses produced in the presence of HsA3F, HsA3G, all three of the artiodactyl A3F proteins or MmA3 showed between 30- and 80-fold more base substitution mutations, which were almost exclusively plus strand G $\rightarrow$ A transition mutations (Figure 5B and Supplementary Figure S3). As described previously by Malim and co-workers, HsA3G with a C-terminal domain E $\rightarrow$ Q mutation failed to cause retroviral hypermutation, although this variant still significantly inhibited HIV-GFP infectivity [Figure 5 and Supplementary Figure S3

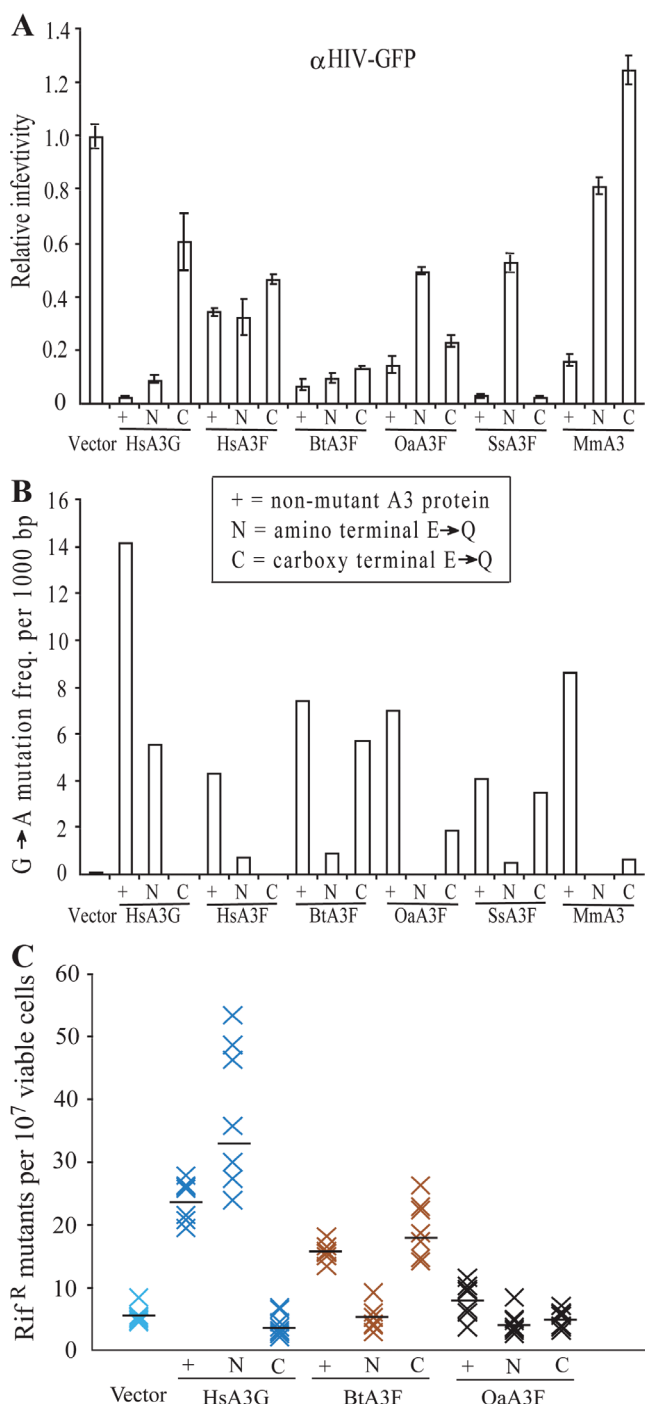


**Figure 4.** Retrovirus restriction activity of the artiodactyl A3F proteins. (A) Infectivity of HIV-GFP produced in the presence of a vector control or the indicated A3 protein. Data were normalized to the infectivity of HIV-GFP produced in the presence of a control vector, which was assigned a value of one. The mean and the SEM of three independent experiments are shown. HIV-1 Vif is not encoded by the proviral vector nor is it included *in trans* in these experiments (contrast with those shown in Figure 6). (B) Infectivity of MLV-GFP produced in the presence of the indicated constructs. Parameters are identical to those in (A).

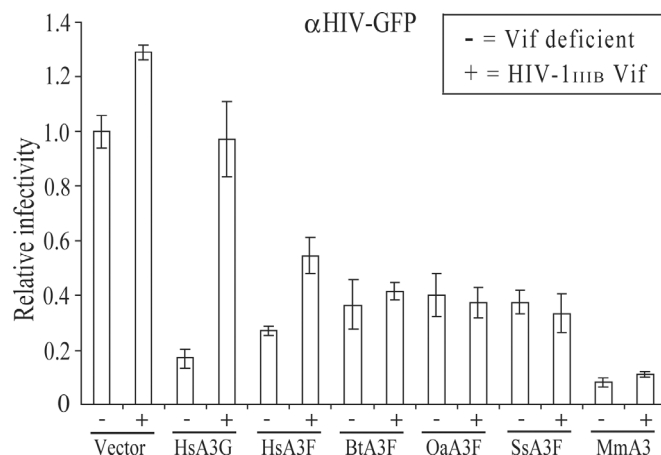
and (32)]. The HsA3F C-terminal Z domain mutant was still able to modestly inhibit HIV-GFP infectivity, without obvious signs of retroviral hypermutation. Interestingly, E→Q substitutions in the N-terminal (but not the C-terminal) domain of all three of the artiodactyl A3F proteins abolished the accumulation of retroviral hypermutations (Figure 5B and Supplementary Figure S3). Thus, these data combined to demonstrate that the N-terminal Z domain of the artiodactyl A3F proteins is catalytic and that both deaminase-dependent and independent activities are required for full levels of retrovirus restriction. In support of this conclusion, the N-terminal Z domain glutamate of BtA3F is required for mutator activity in *E.coli*, whereas the C-terminal Z domain glutamate is dispensable (Figure 5C).

### Retroviral hypermutation properties of artiodactyl A3F proteins

As described above, the *rpoB* mutation spectra of BtA3F, OaA3F and SsA3F suggested that these proteins would trigger retroviral hypermutation patterns biased toward



**Figure 5.** Relative contributions of the N- and C-terminal zinc-binding domains to HIV-GFP restriction. (A) Infectivity of HIV-GFP produced in the presence of a vector control or the indicated non-mutant (+) or mutant A3 protein containing an N-terminal E→Q substitution or a C-terminal E→Q substitution. Parameters are identical to those in Figure 4A. HIV-1 Vif is not included in these experiments (contrast with Figure 6). (B) Frequencies of plus strand retroviral G→A hypermutation observed in HIV-GFP DNA. (C) The N-terminal Z domain glutamate of the indicated artiodactyl A3F protein is required for mutator activity in *E.coli*, whereas the C-terminal Z domain glutamate is largely dispensable. Rif<sup>R</sup> mutation frequencies for eight independent *E.coli* cultures expressing a vector control (light blue), HsA3G (dark blue), BtA3F (brown) or OaA3F (black). The addition sign (+), N and C refer to non-mutant, N-terminal Z domain E→Q mutant or C-terminal Z domain E→Q mutant proteins, respectively. Other parameters are identical to those in Figure 2A.



**Figure 6.** Artiodactyl A3F proteins are resistant to HIV-1 Vif. Parameters are identical to those used in Figure 4A, except for the inclusion of plasmids encoding HIV-1<sub>III<sub>B</sub></sub> Vif (+) or a HIV-1<sub>III<sub>B</sub></sub> ΔVif (–) control, which has translation stop codons at 33 and 34 amino acids (79).

5'-YC, 5'-GC, and 5'-RC, respectively (R = A or G; Y = C or T; Figure 2C). To test this prediction, we examined the types of base substitution mutations and the local retroviral cDNA deamination preferences attributable to expression of the artiodactyl A3F proteins (Figure 5B and Supplementary Figures S3 and S4). In terms of the dinucleotide mutation preferences, the base immediately 5' of the targeted cytosine is a crucial target site determinant. HsA3F and HsA3G overwhelmingly preferred 5'-TC (84%) and 5'-CC (84%), respectively, whereas MmA3 preferred 5'-TC (61%) and 5'-CC (29%). Similar dinucleotide preferences were reported previously for these proteins [e.g. (13,18,51,64,67)]. Roughly paralleling the *E. coli* Rif<sup>r</sup> mutation data, HIV-GFP sequences revealed that the cow and the sheep A3F proteins preferred a pyrimidine (Y) 5' of the deaminated cytosine (93 and 79%, respectively; Supplementary Figure S4). Similarly, the pig A3F protein preferred 5'-GC (47%). This is notable because this is the only example of an A3 protein preferring 5'-purine-C [the immunoglobulin gene deaminase AID also has this preference; e.g. (68)]. In addition, the artiodactyl A3F proteins preferred a T at the –2 position, which is more similar to HsA3F than HsA3G (which prefers a C at this position).

### The artiodactyl A3F proteins are fully resistant to HIV-1 Vif

HsA3F, HsA3G and chimpanzee A3G are neutralized by HIV-1 Vif [e.g. (51,64)]. However, many other monkey A3G proteins (e.g. African green monkey) and MmA3 are completely resistant (27,61–63,69). The full sets of interactions that govern the A3-Vif conflict have not been determined, and the artiodactyl A3F proteins are likely to prove useful in this regard. Therefore, HIV-GFP infectivity was monitored in the presence or absence of HIV-1 Vif and human, artiodactyl or mouse A3 proteins. As described previously, expression of HIV-1 Vif neutralized HsA3F and HsA3G (although the former to a lesser extent) and caused a proportional recovery of HIV-GFP infectivity [Figure 6; e.g. (13,51,64)]. Expression of HIV-1 Vif failed

to enhance the infectivity of HIV-GFP produced in the presence of MmA3 or any of the artiodactyl A3F proteins. Thus, the artiodactyl A3F proteins were fully resistant to HIV-1 Vif.

## DISCUSSION

Murine A3 and many of the human A3 proteins are capable of inhibiting the infectivity of a broad number of retroviruses and endogenous retroelements. Here, we have cloned and characterized a double domain A3 protein from three cloven hoof ungulates, the artiodactyls cattle, sheep and pigs. Artiodactyls are positioned between rodents and primates in most mammalian phylogenetic trees. Thus, the artiodactyl A3F proteins can be considered evolutionary intermediates between the rodent and primate enzymes, and the studies presented here contribute to bridging this vast, 90–100 million year gap. The artiodactyl A3F data help delineate a number of conserved activities that define this group of proteins, including DNA cytosine deaminase activity, cytoplasmic subcellular localization and deaminase-dependent and independent retrovirus restriction activities. Moreover, the artiodactyl A3F proteins also help define non-conserved activities, the varying local DNA cytosine deamination preferences and the differential resistances to viral countermeasures, such as HIV-1 Vif.

One of the most intriguing non-conserved features of the artiodactyl A3F proteins is the fact that a DNA cytosine deaminase activity resides in the N-terminal Z domain, and not in the C-terminal Z domain like several of the human double domain A3 proteins. These data highlight the modular nature of the A3 proteins, which can clearly function as single domain proteins, as double domain proteins with a catalytically active C-terminal Z domain or as double domain proteins with a catalytically active N-terminal Z domain (e.g. HsA3H, HsA3G and SsA3F, respectively). We have not excluded the possibility that under some conditions or with other A3 proteins that both the N- and the C-terminal deaminase domains may be catalytically active. The mouse A3 protein offers an additional dimension of intrigue, as both the N- and the C-terminal zinc-binding deaminase domains are required for retrovirus restriction and for DNA cytosine deaminase activity (Figure 5 and Supplementary Figure 2). Prior studies have demonstrated that the N-terminal, zinc-binding domain of HsA3G mediates interactions with both HIV-1 Vif and Gag (70–76). The C-terminal domain of HsA3G dictates DNA cytosine deaminase activity, and both domains appear to contribute to dimerization activity (31,32,65). Thus, the division of activities between the N- and the C-terminal domains suggests that the A3 proteins are rapidly evolving as modular domains centered upon the conserved, zinc-binding motif. Such modularity may very well enable the A3 proteins to associate with each other to form highly adaptable and potent anti-retroelement defenses. For instance, in a multi-A3 protein complex, only one polypeptide needs to associate with the retroelement target in order to direct the restrictive potential of the other member(s) of the complex. Indeed, the first hints of such a combinatorial restriction potential were observed in retroviral hypermutation loads upon HsA3F and HsA3G co-expression (51).

Taken together with previous studies, it is clear that the mammalian A3 proteins can use both deamination-dependent

and -independent mechanisms to block the transmission of retroviruses and retrotransposons [e.g. (9,12,31–34,65,77)]. As discussed above, our studies indicate that the DNA cytosine deaminase activity can reside in either the N- or the C-terminal Z domain and that this domain appears to be the dominant contributor to restriction of HIV-based retrovirus substrates. Similar observations have been made previously for HsA3G (34,65,66). In contrast, other studies have indicated that either Z domain can mediate full (32,77) or partial (16) levels of HIV-1 restriction. These differential results may be attributable to differences in A3 protein expression levels. A resolution to this apparent paradox may be achieved by working-out the mechanism(s) of DNA deamination-independent restriction.

The studies presented here represent the first essential steps toward understanding the retroelement restriction activities of a third major branch of the vertebrate tree (the other two constituting rodents and primates). Ungulates, and specifically the artiodactyls cattle, sheep and pigs, provide humans with a number of benefits from food products to xenotransplantation possibilities. The possible conflict(s) between BtA3F and bovine immunodeficiency virus (BIV) and those between OaA3F and the sheep lentivirus, maedi-visna virus (MVV), will be of particular future interest. Ongoing studies have indicated that Vif-deficient MVV accumulates retroviral G→A hypermutations with an OaA3F-like dinucleotide mutation spectrum, suggesting that at least OaA3F is active *in vivo* [(78); S. Franzdóttir and V. Andrésdóttir, unpublished data]. Artiodactyl and human comparative studies may also contribute to understanding how HIV-1 uses Vif to neutralize human APOBEC3 proteins.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank R. LaRue for assistance with comparisons and nomenclature, T. Smith for providing cDNA clones and assistance with artiodactyl cDNA sequences, N. Martemyanova for providing expert technical assistance, M. Titus for microscopy facilities, N. Somia for sharing valuable reagents, and M. Murtaugh, P. Hackett, T. Smith and Harris laboratory members for helpful discussions. This work was supported by NIH grant AI064046. RSH is a Searle Scholar and a University of Minnesota McKnight Land Grant Professor. SRJ was the 2004–2005 Val Bjornson Icelandic Exchange Scholarship recipient and is supported in part by a grant from the Icelandic Research Fund. M.D.S. and G.H. were supported in part by a 3M Science and Technology Graduate Fellowship and an NSERC Graduate Studentship, respectively. Assistance with flow cytometry was provided by the University of Minnesota Comprehensive Cancer Center Flow Cytometry Core Facility. University of Minnesota Advanced Genetic Analysis Facility assisted with DNA sequencing. Funding to pay the Open Access publication charges for this article was provided by Spring Point Project.

*Conflict of interest statement.* None declared.

## REFERENCES

- Huthoff, H. and Malim, M.H. (2005) Cytidine deamination and resistance to retroviral infection: towards a structural understanding of the APOBEC proteins. *Virology*, **334**, 147–153.
- Harris, R.S. and Liddament, M.T. (2004) Retroviral restriction by APOBEC proteins. *Nature Rev. Immunol.*, **4**, 868–877.
- Cullen, B.R. (2006) Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors. *J. Virol.*, **80**, 1067–1076.
- Bieniasz, P.D. (2004) Intrinsic immunity: a front-line defense against viral attack. *Nature Immunol.*, **5**, 1109–1115.
- Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J. and Navaratnam, N. (2002) An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics*, **79**, 285–296.
- Wedekind, J.E., Dance, G.S., Sowden, M.P. and Smith, H.C. (2003) Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. *Trends Genet.*, **19**, 207–216.
- Conticello, S.G., Thomas, C.J., Petersen-Mahrt, S. and Neuberger, M.S. (2005) Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol. Biol. Evol.*, **22**, 367–377.
- Chen, H., Lilley, C.E., Yu, Q., Lee, D.V., Chou, J., Narvaiza, I., Landau, N.R. and Weitzman, M.D. (2006) APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr. Biol.*, **16**, 480–485.
- Bogerd, H.P., Wiegand, H.L., Doehle, B.P., Lueders, K.K. and Cullen, B.R. (2006) APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res.*, **34**, 89–95.
- Bogerd, H.P., Wiegand, H.L., Hulme, A.E., Garcia-Perez, J.L., O'Shea, K.S., Moran, J.V. and Cullen, B.R. (2006) Maintaining the integrity of the human genome: a cellular inhibitor of LINE-1 and Alu retrotransposition. *Proc. Natl Acad. Sci. USA*, **103**, 8780–8785.
- Yu, Q., Chen, D., Konig, R., Mariani, R., Unutmaz, D. and Landau, N.R. (2004) APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication. *J. Biol. Chem.*, **279**, 53379–53386.
- Stenglein, M.D. and Harris, R.S. (2006) APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J. Biol. Chem.*, **281**, 16837–16841.
- Bishop, K.N., Holmes, R.K., Sheehy, A.M., Davidson, N.O., Cho, S.J. and Malim, M.H. (2004) Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr. Biol.*, **14**, 1392–1396.
- Dang, Y., Wang, X., Esselman, W.J. and Zheng, Y.H. (2006) Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. *J. Virol.*, doi: 10.1128/JVI.01123-06.
- Sheehy, A.M., Gaddis, N.C., Choi, J.D. and Malim, M.H. (2002) Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*, **418**, 646–650.
- Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L. and Trono, D. (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature*, **424**, 99–103.
- Zhang, H., Yang, B., Pomerantz, R.J., Zhang, C., Arunachalam, S.C. and Gao, L. (2003) The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature*, **424**, 94–98.
- Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S. and Malim, M.H. (2003) DNA deamination mediates innate immunity to retroviral infection. *Cell*, **113**, 803–809.
- Schumacher, A.J., Nissley, D.V. and Harris, R.S. (2005) APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. *Proc. Natl Acad. Sci. USA*, **102**, 9854–9859.
- Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A.J., Heidmann, T. and Schwartz, O. (2005) APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature*, **433**, 430–433.
- Esnault, C., Millet, J., Schwartz, O. and Heidmann, T. (2006) Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses. *Nucleic Acids Res.*, **34**, 1522–1531.
- Dutko, J.A., Schafer, A., Kenny, A.E., Cullen, B.R. and Curcio, M.J. (2005) Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases. *Curr. Biol.*, **15**, 661–666.

23. Turelli, P., Mangeat, B., Jost, S., Vianin, S. and Trono, D. (2004) Inhibition of hepatitis B virus replication by APOBEC3G. *Science*, **303**, 1829.
24. Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., Flory, E., Schumann, G.G. and Munk, C. (2006) APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J. Biol. Chem.*, **281**, 22161–22172.
25. Delebecque, F., Suspene, R., Calattini, S., Casartelli, N., Saib, A., Froment, A., Wain-Hobson, S., Gessain, A., Vartanian, J.P. and Schwartz, O. (2006) Restriction of foamy viruses by APOBEC cytidine deaminases. *J. Virol.*, **80**, 605–614.
26. OhAinle, M., Kerns, J.A., Malik, H.S. and Emerman, M. (2006) Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. *J. Virol.*, **80**, 3853–3862.
27. Mariani, R., Chen, D., Schrofelbauer, B., Navarro, F., Konig, R., Bollman, B., Munk, C., Nymark-McMahon, H. and Landau, N.R. (2003) Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell*, **114**, 21–31.
28. Kobayashi, M., Takaori-Kondo, A., Shindo, K., Abudu, A., Fukunaga, K. and Uchiyama, T. (2004) APOBEC3G targets specific virus species. *J. Virol.*, **78**, 8238–8244.
29. Russell, R.A., Wiegand, H.L., Moore, M.D., Schafer, A., McClure, M.O. and Cullen, B.R. (2005) Foamy virus Bet proteins function as novel inhibitors of the APOBEC3 family of innate antiretroviral defense factors. *J. Virol.*, **79**, 8724–8731.
30. Lochelt, M., Romen, F., Bastone, P., Muckenfuss, H., Kirchner, N., Kim, Y.B., Truyen, U., Rosler, U., Battenberg, M., Saib, A. *et al.* (2005) The antiretroviral activity of APOBEC3 is inhibited by the foamy virus accessory Bet protein. *Proc. Natl Acad. Sci. USA*, **102**, 7982–7987.
31. Haché, G., Liddament, M.T. and Harris, R.S. (2005) The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. *J. Biol. Chem.*, **280**, 10920–10924.
32. Newman, E.N., Holmes, R.K., Craig, H.M., Klein, K.C., Lingappa, J.R., Malim, M.H. and Sheehy, A.M. (2005) Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr. Biol.*, **15**, 166–170.
33. Sasada, A., Takaori-Kondo, A., Shirakawa, K., Kobayashi, M., Abudu, A., Hishizawa, M., Imada, K., Tanaka, Y. and Uchiyama, T. (2005) APOBEC3G targets human T-cell leukemia virus type 1. *Retrovirology*, **2**, 32.
34. Navarro, F., Bollman, B., Chen, H., Konig, R., Yu, Q., Chiles, K. and Landau, N.R. (2005) Complementary function of the two catalytic domains of APOBEC3G. *Virology*, **333**, 374–386.
35. Yu, Y., Xiao, Z., Ehrlich, E.S., Yu, X. and Yu, X.F. (2004) Selective assembly of HIV-1 Vif-Cul5-ElonginB-ElonginC E3 ubiquitin ligase complex through a novel SOCS box and upstream cysteines. *Genes Dev.*, **18**, 2867–2872.
36. Mehle, A., Goncalves, J., Santa-Marta, M., McPike, M. and Gabuzda, D. (2004) Phosphorylation of a novel SOCS-box regulates assembly of the HIV-1 Vif-Cul5 complex that promotes APOBEC3G degradation. *Genes Dev.*, **18**, 2861–2866.
37. Doehle, B.P., Schafer, A., Wiegand, H.L., Bogerd, H.P. and Cullen, B.R. (2005) Differential sensitivity of murine leukemia virus to APOBEC3-mediated inhibition is governed by virion exclusion. *J. Virol.*, **79**, 8201–8207.
38. Mahieux, R., Suspene, R., Delebecque, F., Henry, M., Schwartz, O., Wain-Hobson, S. and Vartanian, J.P. (2005) Extensive editing of a small fraction of human T-cell leukemia virus type 1 genomes by four APOBEC3 cytidine deaminases. *J. Gen. Virol.*, **86**, 2489–2494.
39. Abudu, A., Takaori-Kondo, A., Izumi, T., Shirakawa, K., Kobayashi, M., Sasada, A., Fukunaga, K. and Uchiyama, T. (2006) Murine Retrovirus escapes from murine APOBEC3 via two distinct novel mechanisms. *Curr. Biol.*, **16**, 1565–1570.
40. MacDuff, D.A. and Harris, R.S. (2006) Directed DNA deamination by AID/APOBEC3 in immunity. *Curr. Biol.*, **16**, R186–R189.
41. Petersen-Mahrt, S. (2005) DNA deamination in immunity. *Immunol. Rev.*, **203**, 80–97.
42. Longerich, S., Basu, U., Alt, F. and Storb, U. (2006) AID in somatic hypermutation and class switch recombination. *Curr. Opin. Immunol.*, **18**, 164–174.
43. Zhang, J. and Webb, D.M. (2004) Rapid evolution of primate antiviral enzyme APOBEC3G. *Hum. Mol. Genet.*, **13**, 1785–1791.
44. Sawyer, S.L., Emerman, M. and Malik, H.S. (2004) Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS. Biol.*, **2**, E275.
45. Nei, M., Xu, P. and Glazko, G. (2001) Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc. Natl Acad. Sci. USA*, **98**, 2497–2502.
46. Fahrenkrug, S.C., Smith, T.P., Freking, B.A., Cho, J., White, J., Vallet, J., Wise, T., Rohrer, G., Pertea, G., Sultana, R. *et al.* (2002) Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly. *Mamm. Genome*, **13**, 475–478.
47. Smith, T.P., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C. *et al.* (2001) Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle. *Genome Res.*, **11**, 626–630.
48. Harris, R.S., Petersen-Mahrt, S.K. and Neuberger, M.S. (2002) RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell*, **10**, 1247–1253.
49. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
50. Miyoshi, H., Blomer, U., Takahashi, M., Gage, F.H. and Verma, I.M. (1998) Development of a self-inactivating lentivirus vector. *J. Virol.*, **72**, 8150–8157.
51. Liddament, M.T., Brown, W.L., Schumacher, A.J. and Harris, R.S. (2004) APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 *in vivo*. *Curr. Biol.*, **14**, 1385–1391.
52. Blake, J.A., Eppig, J.T., Bult, C.J., Kadin, J.A. and Richardson, J.E. (2006) The Mouse Genome Database (MGD): updates and enhancements. *Nucleic Acids Res.*, **34**, D562–D567.
53. Yu, M. and Irwin, D.M. (1996) Evolution of stomach lysozyme: the pig lysozyme gene. *Mol. Phylogenet. Evol.*, **5**, 298–308.
54. Reza Shariflou, M. and Moran, C. (2000) Conservation within artiodactyls of an AATA interrupt in the IGF-I microsatellite for 19–35 million years. *Mol. Biol. Evol.*, **17**, 665–669.
55. Petersen-Mahrt, S.K., Harris, R.S. and Neuberger, M.S. (2002) AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature*, **418**, 99–103.
56. Langlois, M.A., Beale, R.C., Conticello, S.G. and Neuberger, M.S. (2005) Mutational comparison of the single-domain APOBEC3C and double-domain APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. *Nucleic Acids Res.*, **33**, 1913–1923.
57. Wichroski, M.J., Robb, G.B. and Rana, T.M. (2006) Human retroviral host restriction factors APOBEC3G and APOBEC3F localize to mRNA processing bodies. *PLoS. Pathog.*, **2**, e41.
58. Marin, M., Rose, K.M., Kozak, S.L. and Kabat, D. (2003) HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nature Med.*, **9**, 1398–1403.
59. Ito, S., Nagaoka, H., Shinkura, R., Begum, N., Muramatsu, M., Nakata, M. and Honjo, T. (2004) Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc. Natl Acad. Sci. USA*, **101**, 1975–1980.
60. McBride, K.M., Barreto, V., Ramiro, A.R., Stavropoulos, P. and Nussenzweig, M.C. (2004) Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. *J. Exp. Med.*, **199**, 1235–1244.
61. Bogerd, H.P., Doehle, B.P., Wiegand, H.L. and Cullen, B.R. (2004) A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor. *Proc. Natl Acad. Sci. USA*, **101**, 3770–3774.
62. Schrofelbauer, B., Chen, D. and Landau, N.R. (2004) A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). *Proc. Natl Acad. Sci. USA*, **101**, 3927–3932.
63. Mangeat, B., Turelli, P., Liao, S. and Trono, D. (2004) A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action. *J. Biol. Chem.*, **279**, 14481–14483.
64. Wiegand, H.L., Doehle, B.P., Bogerd, H.P. and Cullen, B.R. (2004) A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. *EMBO J.*, **23**, 2451–2458.
65. Shindo, K., Takaori-Kondo, A., Kobayashi, M., Abudu, A., Fukunaga, K. and Uchiyama, T. (2003) The enzymatic activity of CEM15/Apobec-3G is essential for the regulation of the infectivity of HIV-1 virion but not

- a sole determinant of its antiviral activity. *J. Biol. Chem.*, **278**, 44412–44416.
66. Iwatani, Y., Takeuchi, H., Strebel, K. and Levin, J.G. (2006) Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. *J. Virol.*, **80**, 5992–6002.
67. Yu, Q., Konig, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J.M. and Landau, N.R. (2004) Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nature Struct. Mol. Biol.*, **11**, 435–442.
68. Beale, R.C., Petersen-Mahrt, S.K., Watt, I.N., Harris, R.S., Rada, C. and Neuberger, M.S. (2004) Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra *in vivo*. *J. Mol. Biol.*, **337**, 585–596.
69. Xu, H., Svarovskaia, E.S., Barr, R., Zhang, Y., Khan, M.A., Strebel, K. and Pathak, V.K. (2004) A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion. *Proc. Natl Acad. Sci. USA*, **101**, 5652–5657.
70. Conticello, S.G., Harris, R.S. and Neuberger, M.S. (2003) The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr. Biol.*, **13**, 2009–2013.
71. Alce, T.M. and Popik, W. (2004) APOBEC3G is incorporated into virus-like particles by a direct interaction with HIV-1 Gag nucleocapsid protein. *J. Biol. Chem.*, **279**, 34083–34086.
72. Cen, S., Guo, F., Niu, M., Saadatmand, J., Deflassieux, J. and Kleiman, L. (2004) The interaction between HIV-1 Gag and APOBEC3G. *J. Biol. Chem.*, **279**, 33177–33184.
73. Svarovskaia, E.S., Xu, H., Mbisa, J.L., Barr, R., Gorelick, R.J., Ono, A., Freed, E.O., Hu, W.S. and Pathak, V.K. (2004) Human apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G) is incorporated into HIV-1 virions through interactions with viral and nonviral RNAs. *J. Biol. Chem.*, **279**, 35822–35828.
74. Luo, K., Liu, B., Xiao, Z., Yu, Y., Yu, X., Gorelick, R. and Yu, X.F. (2004) Amino-terminal region of the human immunodeficiency virus type 1 nucleocapsid is required for human APOBEC3G packaging. *J. Virol.*, **78**, 11841–11852.
75. Zennou, V., Perez-Caballero, D., Gottlinger, H. and Bieniasz, P.D. (2004) APOBEC3G incorporation into human immunodeficiency virus type 1 particles. *J. Virol.*, **78**, 12058–12061.
76. Schafer, A., Bogerd, H.P. and Cullen, B.R. (2004) Specific packaging of APOBEC3G into HIV-1 virions is mediated by the nucleocapsid domain of the gag polyprotein precursor. *Virology*, **328**, 163–168.
77. Bishop, K.N., Holmes, R.K. and Malim, M.H. (2006) Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. *J. Virol.*, **80**, 8450–8458.
78. Kristbjornsdottir, H.B., Andresdottir, V., Svansson, V., Torsteinsdottir, S., Matthiasdottir, S. and Andresson, O.S. (2004) The vif gene of maedi-visna virus is essential for infectivity *in vivo* and *in vitro*. *Virology*, **318**, 350–359.
79. Simon, J.H., Southerling, T.E., Peterson, J.C., Meyer, B.E. and Malim, M.H. (1995) Complementation of vif-defective human immunodeficiency virus type 1 by primate, but not nonprimate, lentivirus vif genes. *J. Virol.*, **69**, 4166–4172.

## SUPPLEMENTARY DATA for Jonsson et al.

**Figure S1.** A non-canonical protein sequence alignment of the artiodactyl APOBEC3F proteins and the C- and N-terminal halves of human APOBEC3F. Alignments were done with Clustal W software. BtA3F and OaA3F proteins are 83% identical. BtA3F and OaA3F are 59% and 61% identical to SsA3F, respectively. Asterisks indicate identical amino acids. The conserved, zinc-binding deaminase domains are boxed. As shown, note that the N-terminal halves of the artiodactyl proteins have 72 identical residues with the C-terminal half of human A3F (compare with Figure S2).

**Figure S2.** A canonical protein sequence alignment of the artiodactyl APOBEC3F proteins and human APOBEC3F. The alignment parameters were similar to those used in Figure S1. As shown, note that the N-terminal halves of the artiodactyl proteins have only 63 identical residues with the N-terminal half of human A3F (compare with Figure S1).

**Figure S3.** Tables summarizing the types of base substitution mutations observed in HIV-GFP. The total number of bases sequenced and the overall percentages of G → A transitions are indicated below each profile. All mutations are designated using the conventional plus-strand nomenclature. All of the non-mutant APOBEC3 proteins analyzed were able to trigger extensive strand-specific G/C → T/A hypermutations.

**Figure S4.** Local DNA cytosine deamination preferences. A comparison of the extended local sequence preferences observed in the minus strand of HIV-GFP produced in the presence of HsA3G, HsA3F, artiodactyl APOBEC3F proteins, MmA3 or the indicated E → Q catalytic mutants. The total number of C → T transition mutations that were detected is indicated.

BtA3F	-----MQPAYRGYSQMPWTRDSSEHMARLDPETFYFQFCNLLYANRRNCSYIC	48
OaA3F	-----MPWISDHVARLDPETFYFQFHNLLYAYGRNCSYIC	35
SsA3F	MDPQRLRQWPGPGPASRGYQRPRIRNPEEWFHELSPRTFSFHRNLRFASGRNRSYIC	60
HsA3Fctd	-----MEAMYPHIFYFHFKNLRKAYGRNESWLC	219
	. : * . * * : * * * * * : * *	
BtA3F	YKVERRKYHSRASFDWGVFHNQVYGGTRCHTELRFLSWFHAEKLRPNERYHITWFMSWSP	108
OaA3F	YRVKTKWHRSPVSFDWGVFHNQVYAGTHCHSERRFLSWFCAKKLRPDECYHITWFMSWSP	95
SsA3F	CQVE-----GKNCFQGIQFQNVPPDPCHAECLFLSWFQSWGLSPDEHYVTVFISWSP	115
HsA3Fctd	FTMEVVKHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSP	279
	: : . : * : * * * . * : * * * * * * * : * : * : * * *	
BtA3F	CMKCAKEVADFLGRHQNVTLTSIFTSRLYKFQEEGSRQGLLRSLDQGAHVDIMSYQEFKYC	168
OaA3F	CMKCAELVAGFLGMYQNVTLTSIFTARLYYFQKPQYRKGLLRSLDQGACVDIMSYQEFKYC	155
SsA3F	CCECAAKVAQFLEENRNVSLSLSAARLYYFWKSESREGLRRLSDLGAQVGIMSFQDFQHC	175
HsA3Fctd	CPECAGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGRLSLSQEGASVEIMGYKDFKYC	339
	* : * * * * * * : * : * * * * . : * * * : * * * * * : * : * *	
BtA3F	WKKFVYSQRRPFRPWKKLDRNYQRLVEELEDILGNT----MNLLREVLFKQQFGNQPRVP	224
OaA3F	WKKFVYSQRRPFRPWKKLKRNYQLLAAELEDILGNT----MNLLRETFLKQQFGNQPRVP	211
SsA3F	WNNFVHNLGMPFQPWKKLHKNYQRLVTELKQILRNT----MNLLKENIFIQQFGNQPRVL	231
HsA3Fctd	WENFVYNDDEPFKPKWGLKYNFLFLDSKLQEILE--	373
HsA3Fntd	MKPHFRNTVERMYRDTFSYNFYNR	24
	* : * * : . * : * * * * . * : * : * * : * : * : * : * : *	
BtA3F	APYYRR-KTYLCYQLKQRN-----DLTLDRGCFRNKKQRHAEIRFIDKINSLDLNPSQS	277
OaA3F	PPYYRR-KTYLCYQLKELD-----DLMLDKGCFRNKKQRHAEIRFIDKINSLNPSQS	264
SsA3F	APYYLR-KTYLCYQVKGPD-----DSLDDKGCFQNKKKRHAEIRFIDKINSLNLDQNC	284
HsA3Fntd	PILSRNTVWLCYEVKTKGSPRRLDAKIFRGQVYSQPEHHAEMCFLSWFCGNQLPAYKC	84
	* . . : * * : * * . * : * * . . : : * * : * . : * : *	
BtA3F	YKIICYITWSPCPNCANELVNFITRNNHLKLEIFASRLYFHWIKSFKMGLQDLQAGISV	337
OaA3F	YKIICYITWSPCPNCASELVDFITRNDHLNLQIFASRLYFHWIKPFGRGLHQLQKAGISV	324
SsA3F	YRIICYVTWSPCHNCAKELVDFISNRHHLSQLFASRLYFHWVRCYQRLQRLQAKRVSV	344
HsA3Fntd	FQITWVFSWTPCPDCVAKLAEFLAEHPNVTLTISAARLYYWERDYRRALCRLSQAGARV	144
	: * : * : * : * : * . : * : * : * : * : * : * : * : * : *	
BtA3F	AVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRRRLQRIILTAPI-----	385
OaA3F	AVMTHTEFEDCWEQFVDNQLRPFQPWDKLEQYSASIRRRRLQRIILTAPT-----	372
SsA3F	AVMKGPEFKDCWEKFVDHQGRSFPSWEKLEQYSESISRRLSRILRFANQNNLEDSFRDLR	404
HsA3Fntd	KIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNP-----	191
	: * * * * * : * : * * : * : * : * : * : * : *	
BtA3F	-----	
OaA3F	-----	
SsA3F	LGSPSPSSSRSDSR	418
HsA3Fntd	-----	

BtA3F	-----MQPAYRGYSQMPWTRDSSEHMARLDPETFYFQFCNLLYANRRNCSYIC	48
OaA3F	-----MPWISDHVARLDPETFYFQFHNNLLYAYGRNCSYIC	35
SsA3F	MDPQRLRQWPGPGPASRGGYGQRPRIRNPEEWFHELSPRTFSFHFRNLRFASGRNRSYIC	60
HsA3F	-----MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLC	36
	: . .: ** : * *	
BtA3F	YKVERRKYHSRASFDWGVFHNQVYGGTRCHTELRFLSWFWHAEKLRPNERYHITWFMSWSP	108
OaA3F	YRVKTWKHRSPVSFDWGVFHNQVYAGTHCHSERRFLSWFCAKKLRPDECYHITWFMSWSP	95
SsA3F	CQVE----GKNCFQGIQFNQVPPDPCHAECLFLSWFQSWGLSPDEHYVYVTFISWSP	115
HsA3F	YEVKTKG-PSRPRLDAKIFRGQVYSQPEHHAEMCFLSWFCGNQLPAYKCFQITWVFSWTP	95
	.*: . : :*:** . *: * ***** . * . : : :*:**:*	
BtA3F	CMKCAKEVADFLGRHQNVTLTSLFTSRLYKFQEEGSRQGLLRLSDQGAHVDIMSYQEFKYC	168
OaA3F	CMKCAELVAGFLGMYQNVTLTSLFTARLYYFQKPQYRKGLLRLSDQACVDIMSYQEFKYC	155
SsA3F	CCECAAKVAQFLEENRVSLSLSAARLYYFWKSESREGLRRLSDLGAQVGIMSFQDFQHC	175
HsA3F	CPDCVAKLAFLAEHPNVTLTISAARLYYWERDYRRALCRLSQAGARVKIMDDEEFAYC	155
	* .*. : * ** **:*: : :*: : : *..* *: * * * * . : * : *	
BtA3F	WKKFVYSQRRPFRPWKKLDRNYQRLVEELEDILGNTMNLRLREVLFKQQFGNQPRVPAPYY	228
OaA3F	WKKFVYSQRRPFRPWKKLDRNYQLLAAELEDILGNTMNLRLRETFLKQQFGNQPRVPPYY	215
SsA3F	WNNFVHNLGMPFQPWKKLHKNYQRLVTELKQILRNTMNLRLKENIFIQQFGNQPRVLAPYY	235
HsA3F	WENFYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEMYPHIFYFHFKNLRKAYGRNE	215
	*: ** : . ** * * : : * : * * : *	
BtA3F	RRKTYLCYQLKQRNDLTLDRCGFRNKKQ---RHAEIRFIDKINSLDLNPQSYYKIICYI	284
OaA3F	RRKTYLCYQLKELDDMLDKGCFRNNKKQ---RHAEIRFIDKINSLNLPQSYYKIICYI	271
SsA3F	LRKTYLCYQVKGPDSDLDKGCFQNNKK---RHAEIRFIDKINSLNLDQNCYRIICYV	291
HsA3F	SWLCFTMEVVKHHSPPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYT	275
	: : * . : : * * : *	
BtA3F	TWSPCPNCANELVNFITRNNHLKLEIFASRLYFHWIKSFKMGLQDLQONAGISVAVMTHTE	344
OaA3F	TWSPCPNCASELVDFITRNDHLNLQIFASRLYFHWIKPFCRGLHQLQKAGISVAVMTHTE	331
SsA3F	TWSPCHNCASELVDFISNRHLSLQLFASRLYFHWVRCYQRLQRLQAKRVSVAVMKGPE	351
HsA3F	SWSPCEPCAGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGRLSLSQEGASVEIMGYKD	335
	: **** : * * : : * : : : * : : * : * : *	
BtA3F	FEDCWEQFVDNQSRPFQPWDKLEQYSASIRRLQRILTAPI-----	385
OaA3F	FEDCWEQFVDNQLRPFQPWDKLEQYSASIRRLQRILTAPT-----	372
SsA3F	FKDCWEKFVDHQGRSFPSWEKLEQYSESISRRLSRILRFANQNNLEDSFRDLRLGSPSPS	411
HsA3F	FKYCWENFVYNDDEPFKPKWGLKYNFLFLDSKLQEILE-----	373
	*: ***: ** : : . * . * : : : * : *	
BtA3F	-----	
OaA3F	-----	
SsA3F	SSRSDSR 418	
HsA3F	-----	

<b>HsA3G</b>				
	A	C	G	T
A		0	0	0
C	0		0	3
G	105	0		0
T	0	0	0	
7380bp, 97.2% G_A				
<b>HsA3G E67Q</b>				
	A	C	G	T
A		0	0	1
C	0		0	0
G	33	0		0
T	0	0	0	
5904bp, 97.1% G_A				
<b>HsA3G E259Q</b>				
	A	C	G	T
A		0	0	0
C	0		0	0
G	0	0		0
T	0	0	0	
5904bp, no mutations				
<b>HsA3F</b>				
	A	C	G	T
A		0	0	0
C	0		0	1
G	32	0		0
T	0	0	0	
7380bp, 97% G_A				
<b>HsA3F E67Q</b>				
	A	C	G	T
A		0	0	0
C	0		0	0
G	5	0		0
T	1	0	0	
6642bp, 83.3% G_A				
<b>HsA3F E251Q</b>				
	A	C	G	T
A		0	0	0
C	0		0	0
G	0	0		0
T	0	1	0	
5904bp, 0% G_A				
<b>BtA3F</b>				
	A	C	G	T
A		0	0	1
C	0		0	2
G	59	0		0
T	0	0	0	
7380bp, 95.2% G_A				
<b>BtA3F E80Q</b>				
	A	C	G	T
A		0	0	0
C	0		0	0
G	2	0		0
T	0	0	0	
2214bp, 100% G_A				
<b>BtA3F E260Q</b>				
	A	C	G	T
A		0	0	0
C	0		0	0
G	17	0		0
T	0	0	0	
2952bp, 100% G_A				
<b>OaA3F</b>				
	A	C	G	T
A		0	0	0
C	0		0	2
G	53	0		1
T	0	0	0	
7380bp, 94.6% G_A				
<b>OaA3F E67Q</b>				
	A	C	G	T
A		0	0	0
C	0		0	0
G	0	0		0
T	0	0	0	
3690bp, no mutations				
<b>OaA3F E247Q</b>				
	A	C	G	T
A		0	0	0
C	0		0	0
G	7	0		0
T	0	0	0	
3690bp, 100% G_A				
<b>SsA3F</b>				
	A	C	G	T
A		0	0	0
C	0		0	0
G	30	0		0
T	0	0	0	
7380bp, 100% G_A				
<b>SsA3F E87Q</b>				
	A	C	G	T
A		0	0	0
C	0		0	0
G	3	0		0
T	0	0	0	
5904bp, 100% G_A				
<b>SsA3F E267Q</b>				
	A	C	G	T
A		0	0	0
C	0		0	0
G	18	0		0
T	0	0	0	
5166bp, 100% G_A				
<b>MmA3</b>				
	A	C	G	T
A		0	0	0
C	0		0	0
G	64	0		0
T	0	0	0	
7380bp, 100% G_A				
<b>MmA3 E73Q</b>				
	A	C	G	T
A		0	0	0
C	0		0	0
G	0	0		0
T	0	0	0	
5904 bp, no mutations				
<b>MmA3 E290Q</b>				
	A	C	G	T
A		0	0	0
C	0		0	0
G	4	0		0
T	0	0	0	
5904bp, 100% G_A				

Jonsson *et al.* **Figure S3**

<b>HsA3G n=105</b>					
nt	-2	-1	C	+1	+2
A	10	10		32	12
C	54	84	<b>100</b>	24	19
G	3	1		22	43
T	33	15		21	26

<b>HsA3G E67Q n=33</b>					
nt	-2	-1	C	+1	+2
A	19	0		39	10
C	42	87	<b>100</b>	19	16
G	16	0		19	39
T	23	13		23	35

<b>HsA3F n=32</b>					
nt	-2	-1	C	+1	+2
A	6	0		45	10
C	16	3	<b>100</b>	13	48
G	16	13		13	16
T	61	84		29	26

<b>HsA3F E67Q n=4</b>					
nt	-2	-1	C	+1	+2
A	25	0		50	50
C	75	0	<b>100</b>	0	0
G	0	50		0	25
T	0	50		50	25

<b>BtA3F n=59</b>					
nt	-2	-1	C	+1	+2
A	13	2		35	4
C	9	62	<b>100</b>	22	33
G	4	5		14	44
T	75	31		38	20

<b>BtA3F E80Q N=2</b>					
nt	-2	-1	C	+1	+2
A	0	0		50	0
C	50	50	<b>100</b>	0	50
G	50	50		0	50
T	0	0		50	0

<b>BtA3F E260Q n=17</b>					
nt	-2	-1	C	+1	+2
A	6	6		47	6
C	12	71	<b>100</b>	24	24
G	0	0		6	41
T	82	24		24	29

<b>OaA3F n=53</b>					
nt	-2	-1	C	+1	+2
A	12	6		37	2
C	10	35	<b>100</b>	21	54
G	0	15		4	31
T	78	44		38	13

<b>OaA3F E247Q n=7</b>					
nt	-2	-1	C	+1	+2
A	14	14		57	29
C	0	43	<b>100</b>	29	29
G	29	0		0	43
T	57	43		14	0

<b>SsA3F n=30</b>					
nt	-2	-1	C	+1	+2
A	23	10		37	7
C	17	27	<b>100</b>	23	30
G	3	47		20	40
T	57	17		20	23

<b>SsA3F E87Q n=6</b>					
nt	-2	-1	C	+1	+2
A	17	50		17	17
C	33	33	<b>100</b>	17	17
G	0	0		33	33
T	50	17		33	33

<b>SsA3F E267Q n=18</b>					
nt	-2	-1	C	+1	+2
A	33	0		33	0
C	27	22	<b>100</b>	22	44
G	22	33		11	0
T	17	44		33	56

<b>MmA3 n=64</b>					
nt	-2	-1	C	+1	+2
A	7	3		53	3
C	9	29	<b>100</b>	14	34
G	0	8		11	47
T	84	61		22	16

<b>MmA3 E290Q n=4</b>					
nt	-2	-1	C	+1	+2
A	0	0		75	0
C	25	25	<b>100</b>	25	25
G	0	0		0	75
T	75	75		0	0

## Paper III



Research article

Open Access

## The artiodactyl *APOBEC3* innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals

Rebecca S LaRue<sup>1</sup>, Stefán R Jónsson<sup>1,2</sup>, Kevin AT Silverstein<sup>3</sup>, Mathieu Lajoie<sup>4</sup>, Denis Bertrand<sup>4</sup>, Nadia El-Mabrouk<sup>4</sup>, Isidro Hötzel<sup>5</sup>, Valgerdur Andrésdóttir<sup>2</sup>, Timothy PL Smith<sup>6</sup> and Reuben S Harris<sup>\*1</sup>

Address: <sup>1</sup>Department of Biochemistry, Molecular Biology and Biophysics, Institute for Molecular Virology, Beckman Center for Genome Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA, <sup>2</sup>University of Iceland, Institute for Experimental Pathology, Keldur v/Vesturlandsveg, 112 Reykjavík, Iceland, <sup>3</sup>Masonic Cancer Center, Biostatistics and Bioinformatics Group, University of Minnesota, Minneapolis, Minnesota 55455, USA, <sup>4</sup>DIRO, Université de Montréal, Montréal, Québec, H3C 3J7, Canada, <sup>5</sup>Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, Washington 99164-7040, USA and <sup>6</sup>USDA/ARS US Meat Animal Research Center, Genetics and Breeding Research Unit, PO Box 166, Clay Center, Nebraska 68933-0166, USA

Email: Rebecca S LaRue - [larue005@umn.edu](mailto:larue005@umn.edu); Stefán R Jónsson - [stefanjo@hi.is](mailto:stefanjo@hi.is); Kevin AT Silverstein - [silve023@umn.edu](mailto:silve023@umn.edu); Mathieu Lajoie - [mathieu.lajoie@gmail.com](mailto:mathieu.lajoie@gmail.com); Denis Bertrand - [dbertran78@gmail.com](mailto:dbertran78@gmail.com); Nadia El-Mabrouk - [mabrouk@iro.umontreal.ca](mailto:mabrouk@iro.umontreal.ca); Isidro Hötzel - [ihotzel@mac.com](mailto:ihotzel@mac.com); Valgerdur Andrésdóttir - [valand@hi.is](mailto:valand@hi.is); Timothy PL Smith - [Tim.Smith@ars.usda.gov](mailto:Tim.Smith@ars.usda.gov); Reuben S Harris<sup>\*</sup> - [rsh@umn.edu](mailto:rsh@umn.edu)

<sup>\*</sup> Corresponding author

Published: 18 November 2008

Received: 13 August 2008

BMC Molecular Biology 2008, 9:104 doi:10.1186/1471-2199-9-104

Accepted: 18 November 2008

This article is available from: <http://www.biomedcentral.com/1471-2199/9/104>

© 2008 LaRue et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** APOBEC3 (A3) proteins deaminate DNA cytosines and block the replication of retroviruses and retrotransposons. Each A3 gene encodes a protein with one or two conserved zinc-coordinating motifs (Z1, Z2 or Z3). The presence of one A3 gene in mice (Z2–Z3) and seven in humans, A3A–H (Z1a, Z2a–Z1b, Z2b, Z2c–Z2d, Z2e–Z2f, Z2g–Z1c, Z3), suggests extraordinary evolutionary flexibility. To gain insights into the mechanism and timing of A3 gene expansion and into the functional modularity of these genes, we analyzed the genomic sequences, expressed cDNAs and activities of the full A3 repertoire of three artiodactyl lineages: sheep, cattle and pigs.

**Results:** Sheep and cattle have three A3 genes, A3Z1, A3Z2 and A3Z3, whereas pigs only have two, A3Z2 and A3Z3. A comparison between domestic and wild pigs indicated that A3Z1 was deleted in the pig lineage. In all three species, read-through transcription and alternative splicing also produced a catalytically active double domain A3Z2–Z3 protein that had a distinct cytoplasmic localization. Thus, the three A3 genes of sheep and cattle encode four conserved and active proteins. These data, together with phylogenetic analyses, indicated that a similar, functionally modular A3 repertoire existed in the common ancestor of artiodactyls and primates (i.e., the ancestor of placental mammals). This mammalian ancestor therefore possessed the minimal A3 gene set, Z1–Z2–Z3, required to evolve through a remarkable series of eight recombination events into the present day eleven Z domain human repertoire.

**Conclusion:** The dynamic recombination-filled history of the mammalian A3 genes is consistent with the modular nature of the locus and a model in which most of these events (especially the expansions) were selected by ancient pathogenic retrovirus infections.

## Background

Mammalian APOBEC3 (A3) proteins have the capacity to potently inhibit the replication of a diverse set of reverse-transcribing mobile genetic elements [1-5]. Susceptible exogenous retroelements include lentiviruses (HIV-1, HIV-2, several strains of SIV and FIV), alpharetroviruses (RSV), betaretroviruses (MPMV), gammaretroviruses (MLV), deltaretroviruses (HTLV), foamy viruses and the hepadnavirus HBV (*e.g.*, [6-14]). Susceptible endogenous retroelements include the yeast retrotransposons Ty1 and Ty2, the murine endogenous retroviruses MusD and Pmv, the murine intracisternal A particle (IAP), the porcine endogenous retrovirus PERV and, potentially, extinct elements such as chimpanzee PtERV1 and human HERV-K, all of which require long-terminal repeats (LTRs) for replication [15-23]. In addition, some A3 proteins can also inhibit L1 and its obligate parasite Alu, retrotransposons that replicate by integration-primed reverse transcription [24-30]. An overall theme is emerging in which most – if not all – retroelements can be inhibited by at least one A3 protein.

However, it is now equally clear that the retroelements of any given species have evolved mechanisms to evade restriction by their host's A3 protein(s). For instance, HIV and SIV use Vif to trigger a ubiquitin-dependent degradation mechanism, foamy viruses use a protein called Bet for an imprecisely defined inhibitory mechanism and some viruses such as MPMV, HTLV and MLV appear to employ a simple avoidance mechanism (*e.g.*, [6,31-34]). Thus, it appears that all 'successful' retroelements have evolved strategies to resist restriction by the A3 proteins of their hosts.

The defining feature of the A3 family of proteins is a conserved zinc(Z)-coordinating DNA cytosine deaminase motif, H-x<sub>1</sub>-E-x<sub>25-31</sub>-C-x<sub>2-4</sub>-C (x indicates a non-conserved position [35,36]). The A3 Z domains can be grouped into one of three distinct phylogenetic clusters – Z1, Z2 or Z3. (Figure 1 & Additional File 1). The Z-based classification system, proposed originally by Conticello and coworkers [35], was revised recently through a collaborative effort [37]. From hereon, the new A3 nomenclature system will be used. Z1 and Z2 proteins have a  $\Delta$ W-S/T-C-x<sub>2-4</sub>-C motif, whereas Z3 proteins have a TW-S-C-x<sub>2</sub>-C motif. Z1 and Z2 proteins can be further distinguished by H-x<sub>1</sub>-E-x<sub>5</sub>-X-V/I and H-x<sub>1</sub>-E-x<sub>5</sub>-W-F motifs, respectively. Z1 proteins also have a unique isoleucine within a conserved RIY motif located C-terminal to the zinc-coordinating residues. At least one protein of each of the Z classes and nearly all identified A3 proteins have exhibited single-strand DNA cytosine deaminase activity. For instance, human A3F, A3G and A3H possess catalytically competent Z2, Z1 and Z3 domains, respectively (*e.g.*, [38-41]).

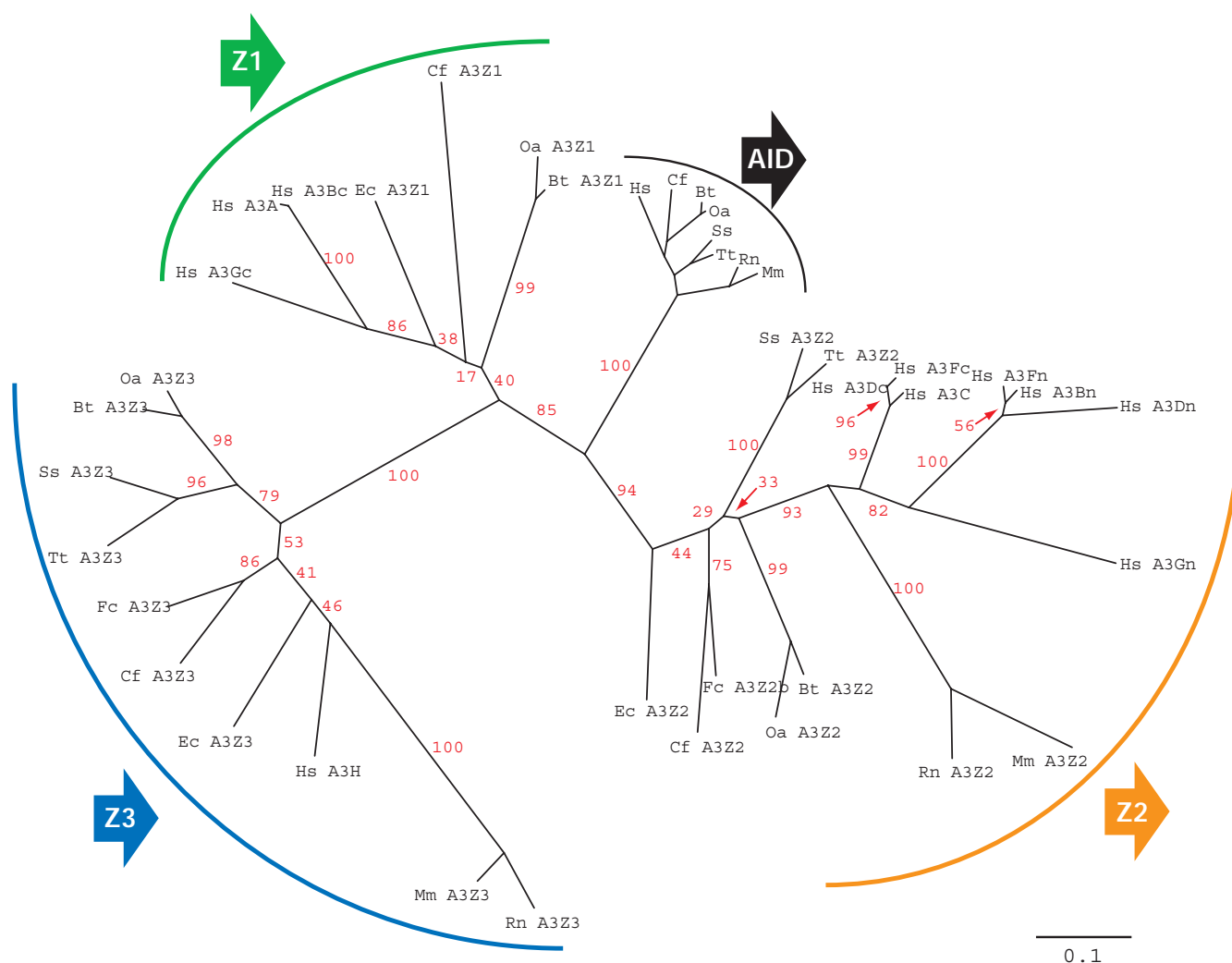
We previously reported a double-domain A3Z2-Z3 gene (formerly called A3F) from the artiodactyls, sheep (*Ovis aries*), cattle (*Bos taurus*) and pigs (*Sus scrofa*) [42]. However, the fact that mammals have varying numbers of A3 genes (*e.g.*, 7 in humans and only 1 in mice) led us to wonder whether additional A3 genes would be present in artiodactyls. To address this point and to learn more about the evolution and functionality of A3 genes in mammals, we sequenced and characterized the full A3 repertoire of sheep and pigs. Here, we demonstrated that sheep and cattle actually have three A3 genes, A3Z1, A3Z2 and A3Z3, with a conserved potential to encode at least four active and distinct proteins (A3Z1, A3Z2, A3Z3 and A3Z2-Z3). We further showed that porcine lineage has a deletion of the orthologous A3Z1 gene and the capacity to encode only three proteins. These data enabled us to deduce that the common ancestor of artiodactyls and primates possessed an A3 repertoire consisting of three Z domains (Z1, Z2 and Z3). Our data further suggested an evolutionary model in which most of the human A3 gene expansion occurred more than 25 million years ago, during early primate evolution and possibly even associated with pathogen-induced population bottlenecks.

## Results

### Sheep and cattle have three A3 genes with a Z1-Z2-Z3 organization

We previously used degenerate PCR, RACE and database mining to identify a cDNA for sheep A3Z2-Z3 (formerly called A3F; [42]). However, because humans have seven A3 genes and mice have only one, we postulated that artiodactyls such as sheep and cattle might have an intermediate number. To address this possibility unambiguously, we sequenced the entire sheep A3 genomic locus. First, a sheep A3Z2-Z3 cDNA was hybridized to a sheep BAC library to identify corresponding genomic sequence. Second, hybridization-positive BACS were screened by PCR for those that also contain the conserved flanking genes *CBX6* and *CBX7*. One BAC was identified that spanned the entire *CBX6* to *CBX7* region, and it was sheared, subcloned, shotgun sequenced, assembled and analyzed (Methods).

DNA sequence analyses revealed that the sheep genomic locus contained another A3 gene between *CBX6* and A3Z2 (Figure 2A). This gene was called A3Z1, because it had sequence characteristics of a Z1-type A3 protein. We therefore concluded that sheep have three A3 genes and, importantly, that each mammalian A3 Z-type was present. This conclusion was supported by the bovine genome assembly, which was released during the course of our studies and showed that cattle also have a sheep-like, three gene A3 repertoire (Figure 2A; Btau\_4.0 <http://www.hgsc.bcm.tmc.edu/projects/bovine/>). The predicted A3Z1 coding sequences of sheep and cattle are 86% iden-



**Figure 1**

**The mammalian A3 Z domains form three distinct phylogenetic groups.** Bootstrap values are indicated in red. The scale bar represents 0.1 nucleotide changes per codon. See the Methods for details. Abbreviations for mammals: Hs = human, Bt = cow, Oa = sheep, Ss = pig, Tt = peccary, Ec = horse, Cf = dog, Fc = cat, Mm = mouse and Rn = rat. Other abbreviations: n = amino terminal domain and c = carboxy terminal domain.

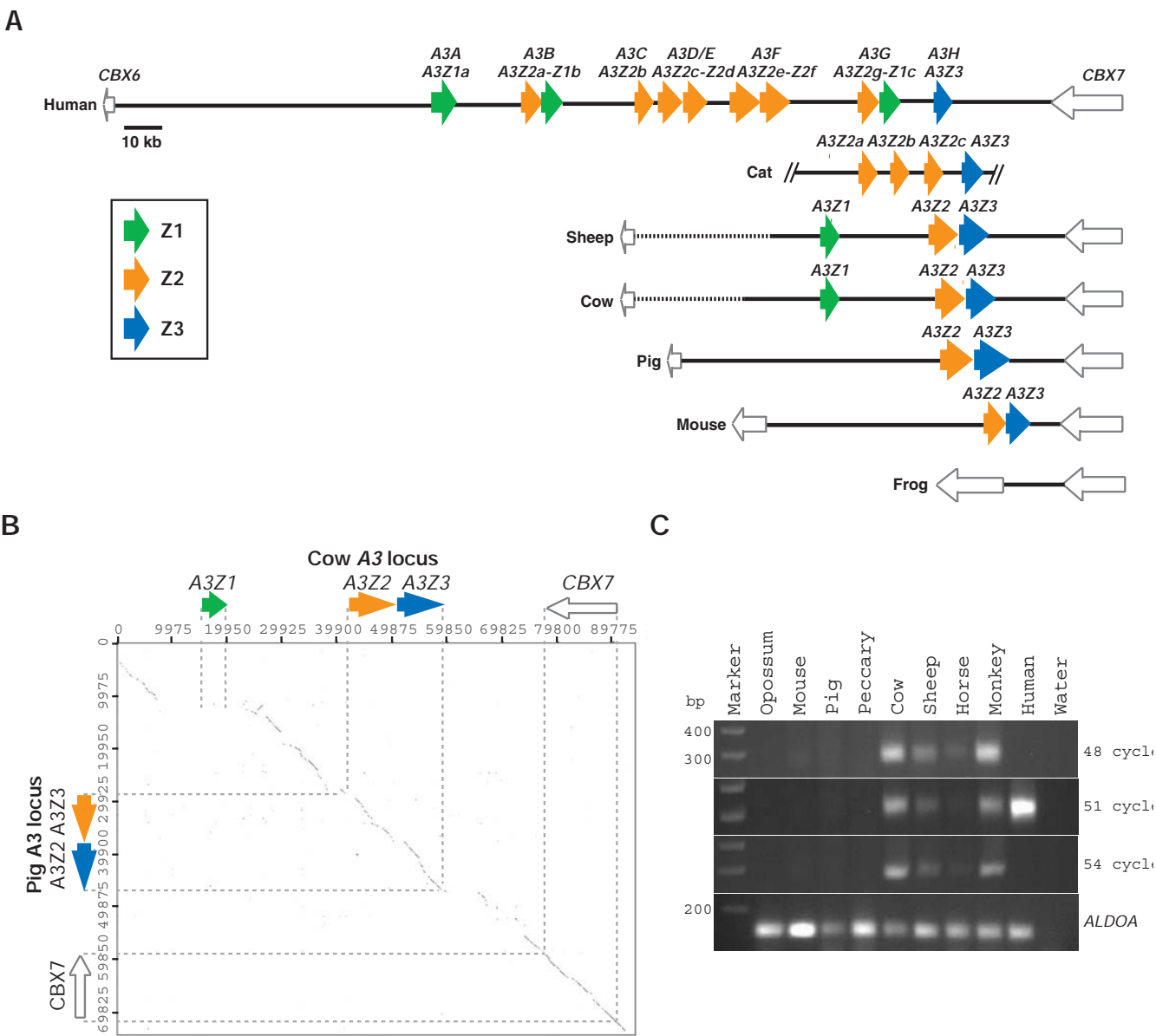
tical, consistent with the fact that these two ruminant artiodactyls shared a common ancestor approximately 14–25 million years ago (MYA) [43–45].

#### **The pig has two A3 genes with a Z2–Z3 organization**

PCR reactions failed to identify an A3Z1-like gene in pigs. Since pigs and cattle/sheep last shared a common ancestor approximately 70–80 MYA [43,44], we considered the possibility that the negative PCR result was not a technical failure and that pigs might actually have a different A3 repertoire. Again, to unambiguously address this possibility, the pig A3 genomic locus was sequenced in entirety. A porcine BAC library was probed with pig A3Z2–Z3 cDNA

and two hybridization-positive BACS were shotgun sequenced. The sequence assemblies revealed that pigs have only two A3 genes A3Z2 and A3Z3 between CBX6 and CBX7 (Figure 2A).

The cattle, sheep and pig A3 locus genomic sequences were compared using dotplot analyses (Figure 2B & Additional File 2). A 22 kb discontinuity was detected between the cow and the pig sequences. The sheep and pig genomic sequences aligned similarly. Multiple (likely inactive) retroelements were found to flank A3Z1 in sheep and cattle. Two were particularly close to the ends of the 22 kb A3Z1 region, a LINE/L1 and a SINE/tRNA-Glu. It is



**Figure 2**  
**The A3 genomic repertoire of sheep, cattle and pigs.** (A) An illustration of the A3 genes of the indicated mammals. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. The conserved flanking genes CBX6 and CBX7 are shown and the scale is indicated. Solid lines represent finished sequence and dotted lines represent gaps or incomplete regions. Non-mammalian vertebrates such as frogs lack A3 genes. (B) A dotplot analysis shows A3Z1 in cattle but not in pig genomic sequence. The x- and y-axis numbers designate nucleotide positions within the indicated genomic consensus sequences. (C) PCR analysis of genomic DNA from the indicated species showing that a 250–256 bp Z1-specific amplicon can only be obtained from a subset of mammals. The human 51 and 54 cycle amplicons were too abundant to be run on the same gel. Monkey genomic DNA is from the African green monkey. The ALDOA gene was used as a positive control (115 bp).

possible that one of these elements mediated a simple direct repeat recombination event that deleted the A3Z1 region in pigs. However, we were unable to identify such a causative retroelement in the pig genomic sequence.

To begin to address whether the potential A3Z1 deletion in pigs occurred recently (e.g., a rare deletion fixed by selective breeding) or whether it was more ancient, we asked whether a non-domesticated, distant relative of the

pig, the collared peccary (*Tayassu tajacu*), has an A3Z1 gene. Lineages leading to present-day domesticated pigs and the peccary diverged approximately 25–35 MYA [43]. A pan-species, A3Z1 PCR primer set was developed and used in these experiments. In contrast to human, African green monkey, horse, cow and sheep genomic DNA which yielded a 250–256 bp Z1-specific PCR products confirmable by DNA sequencing, the genomic DNA of domesticated pig, the collared peccary, mice and opossum failed to yield a product even after 54 cycles (Figure 2C). A highly conserved gene, *ALDOA*, was used as a PCR control to demonstrate the integrity of the genomic DNA samples.

Interestingly, Z1 PCR product sequencing and recently released EST sequences revealed that the related hoofed mammal, the horse, also has a Z1-type A3 gene (Figure 1C & Additional File 3). Two-'toed' hoofed animals such as sheep, cattle and pigs belong to the ungulate order artiodactyla (even-toe number), and one-'toed' hoofed animals such as horses belong to the ungulate order perissodactyla (odd-toe number). Since these two ungulate orders diverged approximately 80–90 MYA [43,44] and both have species with Z1-type A3 genes, it is highly likely that the common ancestor also had an A3Z1 gene (as well as A3Z2 and A3Z3 genes). It is therefore highly unlikely that an A3Z1 gene independently appeared at the same genomic position in artiodactyls, perissodactyls and primates. Rather, all of the data support a model where a common ancestor of the domesticated pig and the collared peccary experienced a 22 kb deletion that resulted in the loss of A3Z1 (*i.e.*, a divergent evolutionary model). Furthermore, since artiodactyls, perissodactyls and humans shared a common ancestor approximately 80–120 MYA [43,44], the presence of Z1-type A3 genes in both the primate and the artiodactyl limbs of the mammalian tree is also most easily explained by common ancestry. Thus, our combined datasets indicated that this ancestor possessed a full A3 Z repertoire, with one of each type of Z domain (Z1, Z2 and Z3), the minimal substrate required to evolve into the present-day eleven Z domain human A3 locus (discussed further below).

### **The artiodactyl A3Z2 and A3Z3 genes combine to encode 3 distinct mRNAs and proteins**

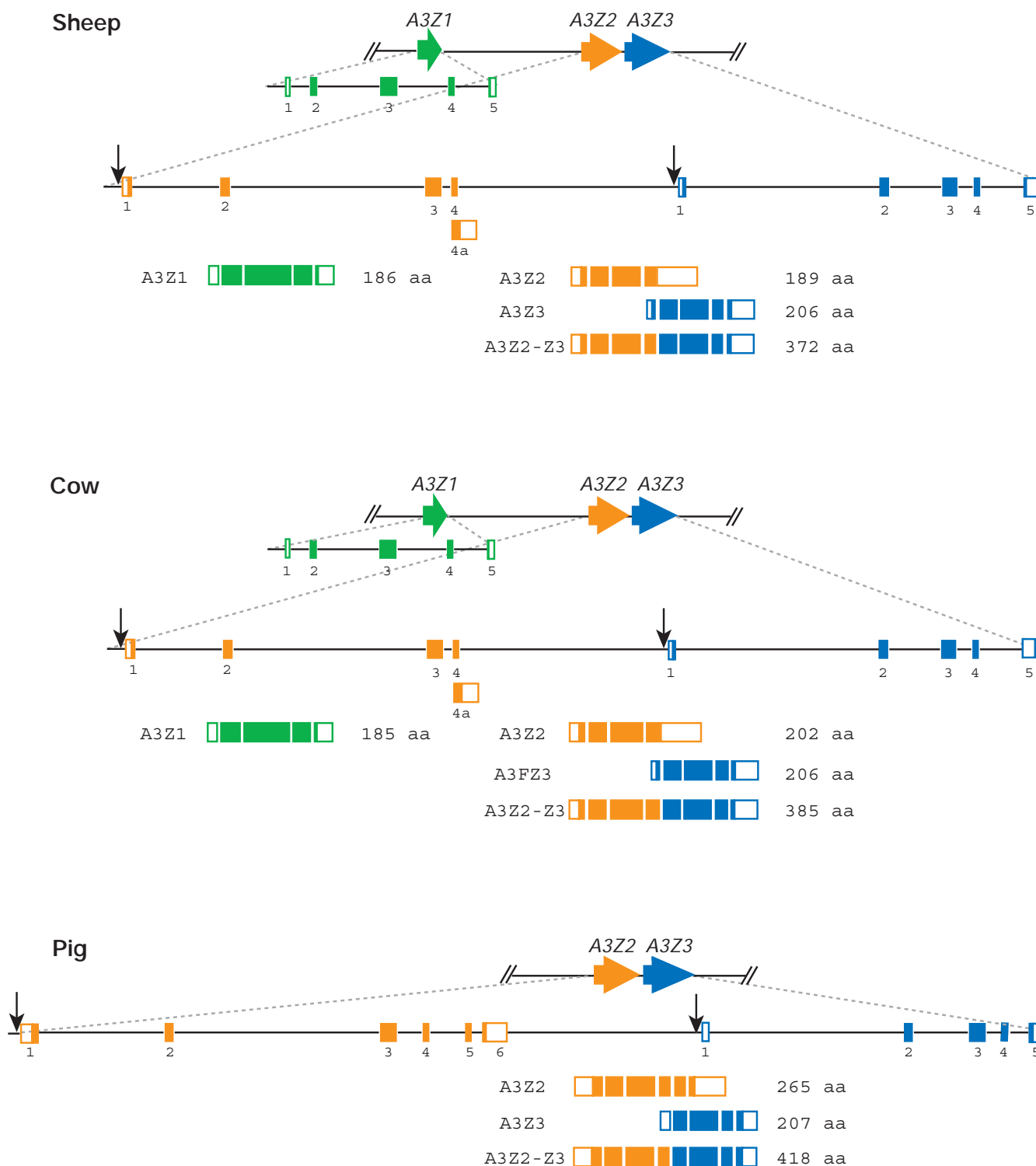
We previously characterized several activities of the double-domain A3Z2-Z3 protein from cattle, sheep and pigs [42]. While re-confirming the 5' and 3' ends of the A3Z2-Z3 transcripts by RACE, we discovered two interesting variants that were conserved between these three species. First, using sheep and cattle PBMC or cell line cDNA (FLK and MDBK, respectively), 3' RACE frequently produced a smaller than expected fragment. The sequence of this fragment indicated the existence of a short 1037 bp transcript due to premature termination 329 or 330 nucleotides into intron 4 for sheep and cattle, respectively (Figure 3). This

truncated transcript was readily amplified from sheep and cattle PBMCs and represented by existing EST sequences (Additional File 3 and data not shown). Therefore, this novel transcript was predicted to result in a single-domain Z2 protein, A3Z2, with a length of 189 and 202 amino acids for sheep and cattle, respectively (Figure 3 & Additional File 3). A pig A3Z2 transcript was also identified by RACE and EST sequences but, in contrast to sheep and cattle, exon 4 was spliced to two additional exons before terminating prematurely (Figure 3 & Additional File 3). As a consequence, pig A3Z2 was predicted to be 265 amino acids. These analyses indicated that artiodactyls have the capacity to express a single domain A3Z2 protein, in contrast to what we had deduced previously [42].

Second, 5' RACE data and cattle and pig EST sequences suggested that yet another mechanism served to broaden the coding potential of the artiodactyl A3 locus (Additional File 3 and data not shown). Several transcripts appeared to originate from the region immediately upstream of A3Z3, whereas our prior studies had only detected transcripts originating upstream of A3Z2 [42]. A comparison of cDNA and genomic sequences revealed the presence of an exon in this location (A3Z3 exon 1 in Figure 3). Transcripts initiating here produced 941 (sheep), 964 (cow) or 1003 (pig) nucleotide messages. The resulting A3Z3 protein was predicted to be 206 residues for sheep and cattle and 207 for pigs (Figure 3).

The A3Z3 mRNA data strongly suggested the existence of an internal promoter. This was supported by cis-regulatory element prediction algorithms, which identified a conserved interferon-stimulated response element (ISRE) upstream of A3Z3, as well as upstream of A3Z2 (Figure 3 & Additional File 4). These ISREs were strikingly similar to those located in the promoter regions of human *A3DE*, *A3F* and *A3G*, supporting the likelihood that interferon-inducibility is a conserved feature of many mammalian A3 genes (*e.g.*, [8,46–49]). These putative ISREs significant similarity to functional elements in known interferon-inducible genes *ISG54* and *ISG15* [50–52]. We also predicted binding sites for another well-known transcription factor, Sp1, upstream of the A3Z3 transcription start site. This activator was also recently reported for human and cat A3 genes ([53,54]; LaRue & Harris, data not shown).

Together with our previous data on the double domain A3 protein of these artiodactyl species, A3Z2-Z3 [42], these expression and promoter data revealed that two single-domain A3 genes can readily encode at least three distinct proteins – A3Z2, A3Z3 and A3Z2-Z3. A similar strategy may also be used by rodents, which also have an A3 gene with Z2 and Z3 domains. A similar modularity was reported recently for the cat A3 locus, where two single domain A3 genes combined to produce a functional dou-

**Figure 3**

**The coding potential of the sheep, cow and pig A3 genes.** Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. The exons are shown below the gene schematics with coding regions represented by filled boxes and untranslated regions by open boxes. The gene schematics and exon blow-ups are drawn to scale. Arrows indicate approximate positions of predicted ISREs (Additional File 4). See the main text and the Methods for details.

ble-domain A3 protein [54]. We suggest that combining single-domain A3s to yield functionally unique double-domain proteins may be a general strategy used by many mammals to bolster their A3-dependent innate immune defenses.

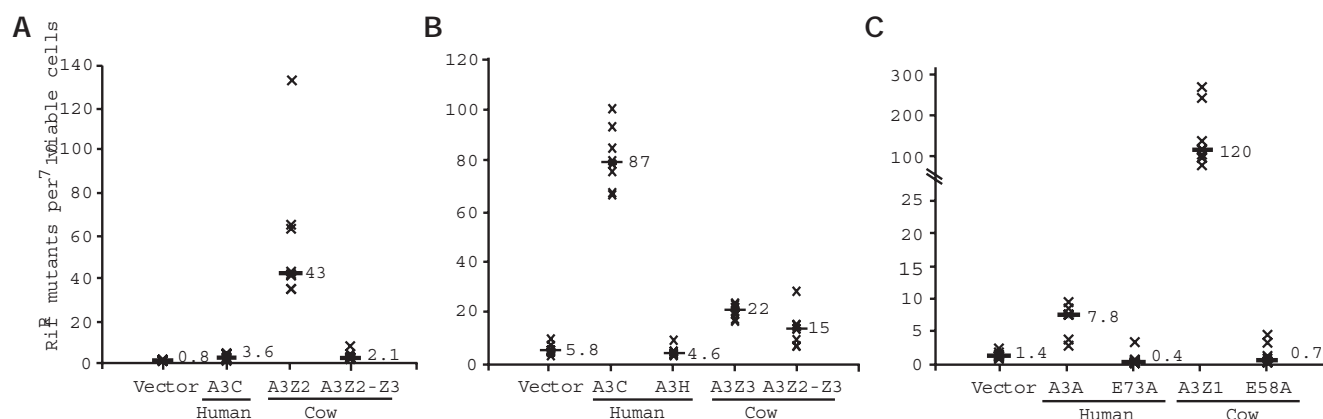
#### All four artiodactyl A3 proteins – A3Z1, A3Z2, A3Z3 and A3Z2-Z3 – elicit DNA cytosine deaminase activity

All currently described A3 proteins have elicited single-strand DNA cytosine to uracil deaminase activity in one or more assays (e.g., [24,41,42,54-59]). For instance, we showed that the artiodactyl A3Z2-Z3 proteins could catalyze the deamination of *E. coli* DNA and retroviral cDNA [42]. However, catalytic mutants indicated that only the N-terminal Z2 domain of cow, sheep and pig A3Z2-Z3 was active. This observation contrasted with data for the double-domain human A3B, A3F and A3G proteins, where the C-terminal domain clearly contains the dominant active site (e.g., [30,38-40,42,60]). Nevertheless, these datasets suggested that the double-domain A3 proteins have separated function, with one domain predominantly serving as a catalytic center and the other as a regulatory center.

However, a recent study with human A3B indicated that both Z domains have the potential to be catalytically active [61]. It was therefore reasonable to ask whether the single domain A3Z2 and A3Z3 proteins of artiodactyls would be capable of DNA cytosine deamination in an *E. coli*-based activity assay. Elevated frequencies of rifampicin-resistance (Rif<sup>R</sup>) mutations in *E. coli* provide a

quantitative measure of the intrinsic A3 protein DNA cytosine deaminase activity (e.g., [38,40,56,57]). In contrast to full-length cow A3Z2-Z3, which triggered a modest 2-fold increase in the median Rif<sup>R</sup> mutation frequency over the vector control, non-induced levels of cow A3Z2 caused a large 50-fold increase (Figure 4A). The pTrc99-based vector used in these studies has an IPTG-inducible promoter, and induced levels of cow A3Z2 prevented *E. coli* growth, presumably through catastrophic levels of DNA cytosine deamination. In contrast, induced levels of sheep or pig A3Z2 proteins were not lethal, but their expression also caused significant increases in the median Rif<sup>R</sup> mutation frequency (Additional File 5 and LaRue & Harris, data not shown). Thus, as anticipated by our prior studies, the A3Z2 proteins of cattle, sheep and pigs showed intrinsic DNA cytosine deaminase activity.

We were therefore surprised that induced levels of the cow single-domain protein A3Z3 also caused a significant 4-fold increase in the median Rif<sup>R</sup> mutation frequency (Figure 4B). This result contrasted with the related Z3 protein of humans, A3H, which appeared inactive in this assay (Figure 4B & Additional File 3). However, it is worth noting that other Z3-type A3 proteins, a different human A3H variant, African green monkey A3H, rhesus macaque A3H and cat A3Z3 (formally A3H), all showed evidence for DNA deaminase activity in the *E. coli*-based mutation assay and/or in retrovirus infectivity assays [41,54,62,63]. Thus, our intended human A3H control appears to be the exception rather than the rule and that the single-domain



**Figure 4**

**The artiodactyl A3 proteins catalyze DNA cytosine deamination.** (A) Cow A3Z2 triggers a strong mutator phenotype in *E. coli*. Rif<sup>R</sup> mutation frequency of 4–6 independent bacteria cultures expressing basal levels of indicated A3 proteins (each X represents data from a single culture). To facilitate comparisons, the median mutation frequency is indicated for each condition. (B) Cow A3Z3 triggers a modest mutator phenotype in *E. coli*. Labels and conditions are similar to those in panel (A), except IPTG was used to induce protein expression. (C) Non-induced cow A3Z1 triggers a strong mutator phenotype in *E. coli*, which is completely abrogated by substituting the catalytic glutamate (E58) for alanine. Labels are similar to those in panel (A).

A3Z3 protein of artiodactyls is capable of DNA cytosine deaminase activity.

We also observed that the artiodactyl A3Z1 protein was capable of robust DNA cytosine deaminase activity (*e.g.*, Figure 4C and LaRue and Harris, unpublished data). This result was fully anticipated based on the fact that the related Z1 domain proteins of humans A3A, A3B and A3G are catalytically active [24,30,61,64]. However, it is worth noting three observations suggesting that cow A3Z1 is the most active of all reported A3 proteins. First, we were never able to directionally clone (even non-induced) A3Z1 of sheep or cattle into pTrc99A, which has a leaky promoter. Second, we were only able to topoisomerase-clone cow A3Z1 in a direction opposite to the *lac* promoter ( $n > 12$ ). Finally, even with cow A3Z1 in the promoter-opposing orientation in the topoisomerase cloning plasmid, we observed 100-fold increases in Rif<sup>R</sup> mutation frequency in the *E. coli*-based mutation assay that were fully dependent on the catalytic glutamate E58A (presumably due to expression from a cryptic promoter; Figure 4C). To summarize this section, all four of the A3 proteins of artiodactyls demonstrated intrinsic DNA cytosine deaminase activity.

#### **A3Z1, A3Z2, A3Z3 and A3Z2-Z3 differentially localize in cells**

Fluorescent microscopy was used to examine the subcellular distribution of each of the artiodactyl A3 proteins fused to GFP. Like the human A3 proteins, which each have unique overall subcellular distributions, we imagined that distinct localization patterns might correlate with differential functions. For instance, the first column of Figure 5 shows representative images of live HeLa cells expressing human A3F-GFP, A3A-GFP, A3C-GFP and A3H-GFP, which predominantly localize to the cytoplasm, cell-wide with a nuclear bias, cell-wide and cell-wide with a clear nucleolar preference, respectively. Cow A3Z1-GFP showed an indiscriminate cell-wide distribution similar to that of human A3A-GFP and GFP alone (Figure 5, second row and data not shown).

As shown previously, cattle and pig A3Z2-Z3-GFP localize to the cytoplasm, with some cells showing bright aggregates (Figure 5, row 1; [19,42]). Cattle and pig A3Z2 also appeared predominantly cytoplasmic, but a significant fraction clearly penetrated the nuclear compartment (row 3). The subcellular distribution of cattle and pig A3Z2 differed from the similarly sized Z2 protein human A3C, which was cell-wide, and it is therefore likely that an active process underlies the cytoplasmic bias of the artiodactyl A3Z2 proteins. Interestingly, the A3Z3 proteins of cattle and sheep, like human A3H, localized cell-wide with clear accumulations in the nucleoli (row 4). Similar data were obtained using these GFP fusion constructs in live cattle MDBK cells and in live pig PK15 cells (LaRue &

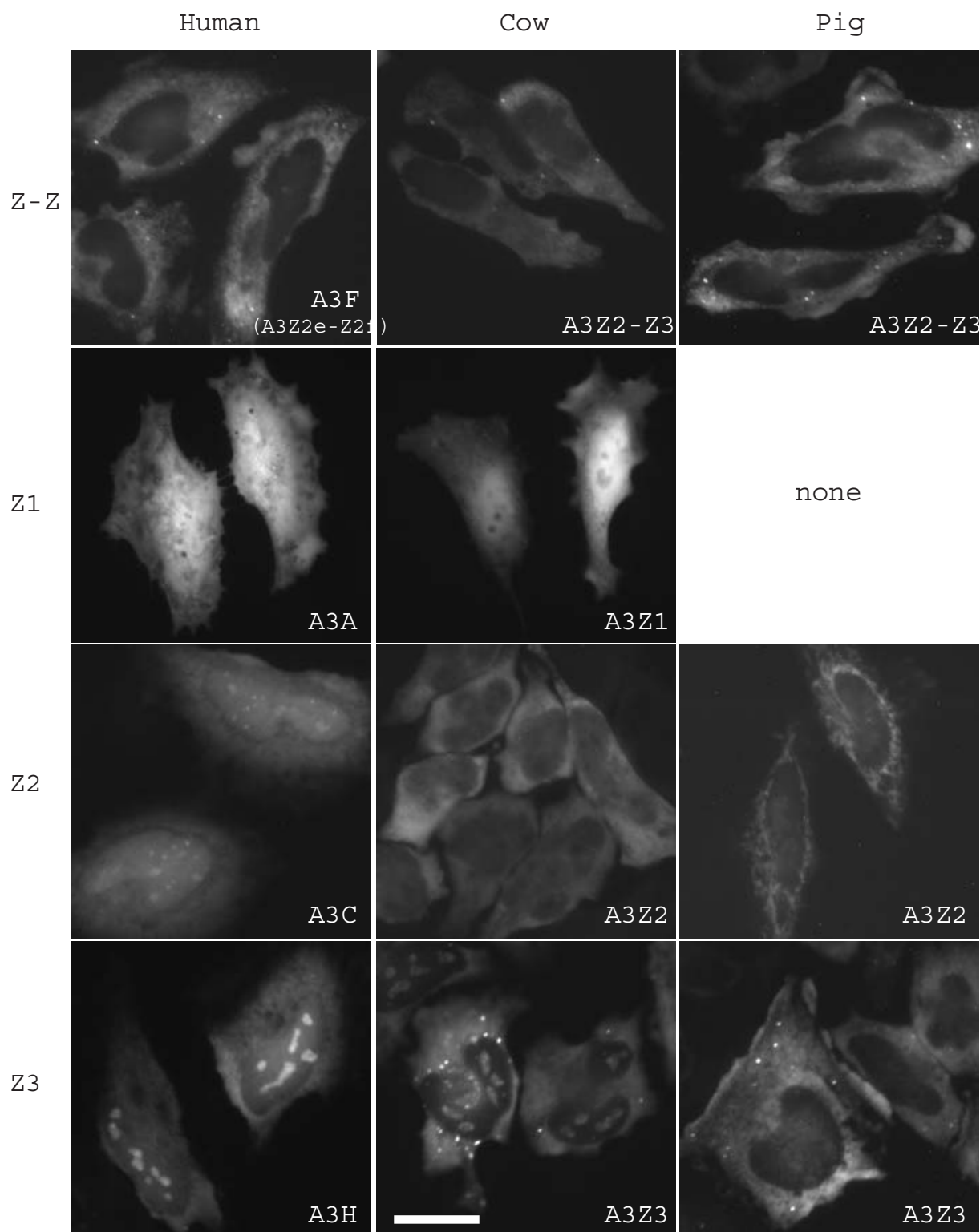
Harris, data not shown). These fluorescent microscopy observations demonstrated that all of the artiodactyl A3 proteins can be expressed in mammalian cells and that they have both distinct and overlapping subcellular distributions.

#### **The artiodactyl A3 genes show evidence for positive selection**

Many human, non-human primate and feline A3 genes show signs of strong positive selection, which can be interpreted as evidence for a history filled with pathogen conflicts [41,54,65,66]. However, given the relative stability of the artiodactyl A3 locus, at least in terms of gene number, we wondered whether the artiodactyl A3 genes might be under less intense selective pressure (perhaps even neutral or negative). This possibility was assessed using two methods to compare the number of mutations that resulted in amino acid replacements to the number that were silent between pairs of artiodactyl species. This ratio of replacement (dN) to silent (dS) mutations yields an omega ( $\omega$ ) value, which if greater than one is indicative of positive selection, if equal to one of neutral selection and if less than one of negative selection. We focused these analyses on the single exon that encodes the conserved Z domain to minimize potentially confounding effects from recombination.

We first generated a combined phylogeny for each distinct A3 Z domain and its inferred ancestral sequences (Additional File 6). Using the PAML free ratio model, the artiodactyl A3Z1 and the A3Z2 genes appeared to be under a weak negative selection pressure, with  $\omega$  values uniformly below one (Additional File 6). Similarly, since the existence of the last common ancestor of cattle and sheep or of the pig and peccary, the artiodactyl A3Z3 genes showed evidence for weak negative selection pressure (Additional File 6C). However, a comparison of the inferred ancestral ruminant sequence with the inferred porcine sequence yielded a  $\omega$  value of 1.5, suggesting that the ancestor(s) of modern day artiodactyls may have experienced intermittent positive selection (Additional File 6C). These values were not as high as those for primate A3Z3 (A3H data originally reported by [41] and re-calculated here with a representative clade shown in Additional File 6C). Moreover, all of these data contrasted sharply with the artiodactyl and primate AID genes, which are under an obvious strong negative selection pressure presumably for essential functions in antibody diversification.

However, because the free ratio model averages all possible sites and has a tendency to underestimate instances of positive selection, we subsequently used PAML NsSites to do a more focussed examination of artiodactyl A3 Z domain variation. Several distinct selection models were used (M2 and M8 and two codon frequency models F61 and F3  $\times$  4), and each yielded significant signs of positive



**Figure 5**  
**The subcellular distribution of cow and pig A3 proteins in comparison to human A3 proteins with similar Z domains.** Representative images of live HeLa cells expressing the indicated A3-GFP fusion proteins are shown. The scale bar represents 25  $\mu$ m.

selection (Table 1; see Methods for procedural details and Additional File 3 for sequence information). The Z3 domain A3 genes of sheep, cattle, pig, peccary and horse showed the highest dN/dS ratios, ranging from 4.4 to 5.8 and indicating that 22–31% of the residues were subjected to positive selection. Lower but still significant positive dN/dS ratios were obtained for the Z2 domain A3 genes (1.7 to 2.3 with 33 to 46% of the residues under positive selection). Moreover, together with available dog and horse Z1 sequences, the Z1 A3 genes of cattle and sheep showed intermediate degrees of positive selection, with dN/dS ratios of 2.5 to 3.9 and 28 to 50% of the residues under some degree of positive selection (Table 1 & Additional File 3). Thus, similar to most other mammals analyzed to date, the artiodactyl A3 genes have been subjected to strong evolutionary pressure (see Discussion).

### A3 Z domain distribution in mammals

Our studies strongly indicated that the present-day A3 locus of sheep and cattle resembles one that existed in the common ancestor of placental mammals, consisting of precisely one of each of the three phylogenetically distinct Z domains: Z1, Z2 and Z3 (Figure 6; also see Figure 1 & Additional File 3). Molecular phylogenetic data helped us infer that such a common ancestor existed approximately 100–115 MYA [43,44]. However, the bulk of the primate A3 gene expansion most likely occurred more recently because the main branches leading to rodents and humans split 90–110 MYA. It is therefore likely that rodents lost a Z1 A3 gene after branching off of the main mammalian tree (like pigs, cats and some humans; see Figure 6 & Discussion). Moreover, the recently published draft of the rhesus macaque genome helped to further whittle-down when the bulk of the primate-specific

expansion occurred, because these animals also possess a human-like A3 gene repertoire (Figure 6; [41,67,68] and our unpublished data). Thus, since the human and macaque lineages diverged approximately 25 MYA [43,67,69], the massive expansion from the inferred sheep/cow-like Z1-Z2-Z3 A3 gene set to a locus resembling the present-day human repertoire must have occurred within a relatively short 65–85 million year period (indicated by an asterisk in Figure 6).

### A minimum of 8 recombination events were required to generate the present-day human A3 locus from the common ancestor of artiodactyls and primates

The inferred ancestral Z1-Z2-Z3 locus was used as a starting point to deduce the most likely evolutionary scenario that transformed it into the much larger eleven Z domain human A3 repertoire. Two types of recombination events were considered, tandem duplications (obviously required for A3 gene expansion) and deletions. Self-similarities in the DNA sequence of the human A3 locus provided strong evidence for prior tandem duplications by unequal crossing-over (for more details on tandem duplication modeling see [70,71]). This mode of evolution is also supported by the fact that the human A3 locus contains many retroelements that could serve as substrates for homologous recombination [35]. Since our present studies showed that the Z domains are highly modular and capable of individual function, they were considered as the core units for duplications in our inference procedures (*i.e.*, an unequal cross-over event can simultaneously duplicate one or more tandemly arranged Z-domains and associated flanking sequences). Similarly, deletions could involve one or more Z domains and result from unequal crossing-over or intra-chromosomal events.

**Table 1: Evidence for positive selection in the artiodactyl Z domains.**

Z domain <sup>a</sup>	Codon frequency model <sup>b</sup>	Comparison of null and positive selection models <sup>c</sup>	Significance	Tree length <sup>d</sup>	dN/dS (%) <sup>e</sup>
Z1	F6I	M1–M2	p = 0.01	2.6	2.5 (50)
		M7–M8	p = 0.01	2.6	2.5 (50)
	F3 × 4	M1–M2	p = 0.04	3.9	3.9 (28)
		M7–M8	p = 0.02	3.9	3.9 (33)
Z2	F6I	M1–M2	p = 0.005	2.4	2.3 (46)
		M7–M8	p = 0.004	2.4	2.3 (45)
	F3 × 4	M1–M2	p = 0.3	3.0	1.7 (27)
		M7–M8	p = 0.04	3.0	1.7 (33)
Z3	F6I	M1–M2	p < 0.001	2.4	4.5 (30)
		M7–M8	p < 0.001	2.4	4.4 (31)
	F3 × 4	M1–M2	p < 0.001	3.1	5.8 (22)
		M7–M8	p < 0.001	3.1	5.7 (23)

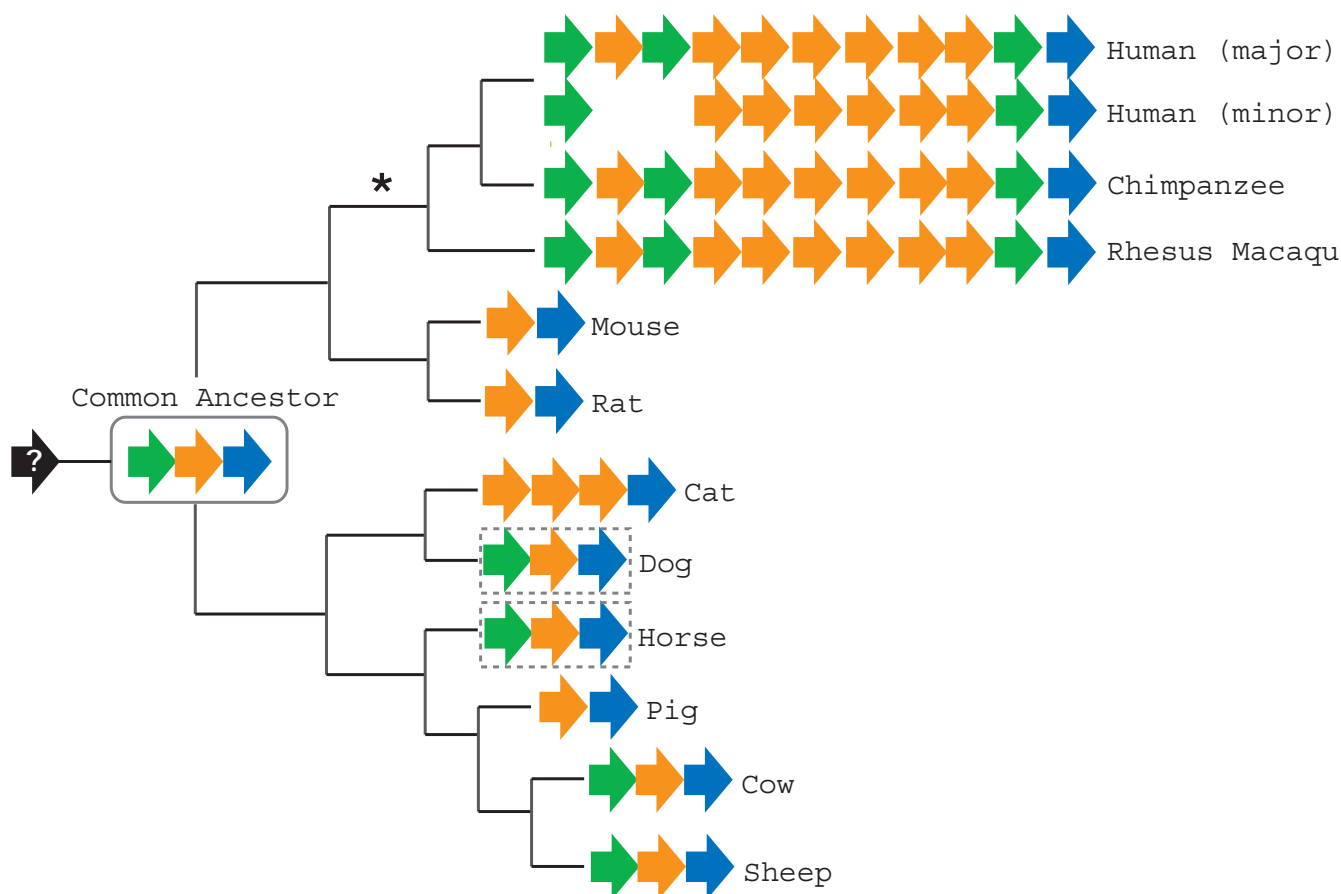
<sup>a</sup>Only the sequences of Z domain-encoding exons were used in these analyses (see Table S1 for GenBank accessions).

<sup>b</sup>Two different codon frequency models were used to minimize potentially artificial results.

<sup>c</sup>Likelihood ratio tests were done to compare the null models M1 and M7 to each positive selection model M2 and M8, respectively, using PAML NsSites.

<sup>d</sup>Tree length provides a measure of nucleotide substitutions per codon along all combined phylogenetic branches.

<sup>e</sup>The percentage of all codons influenced by positive selection is indicated in parentheses.

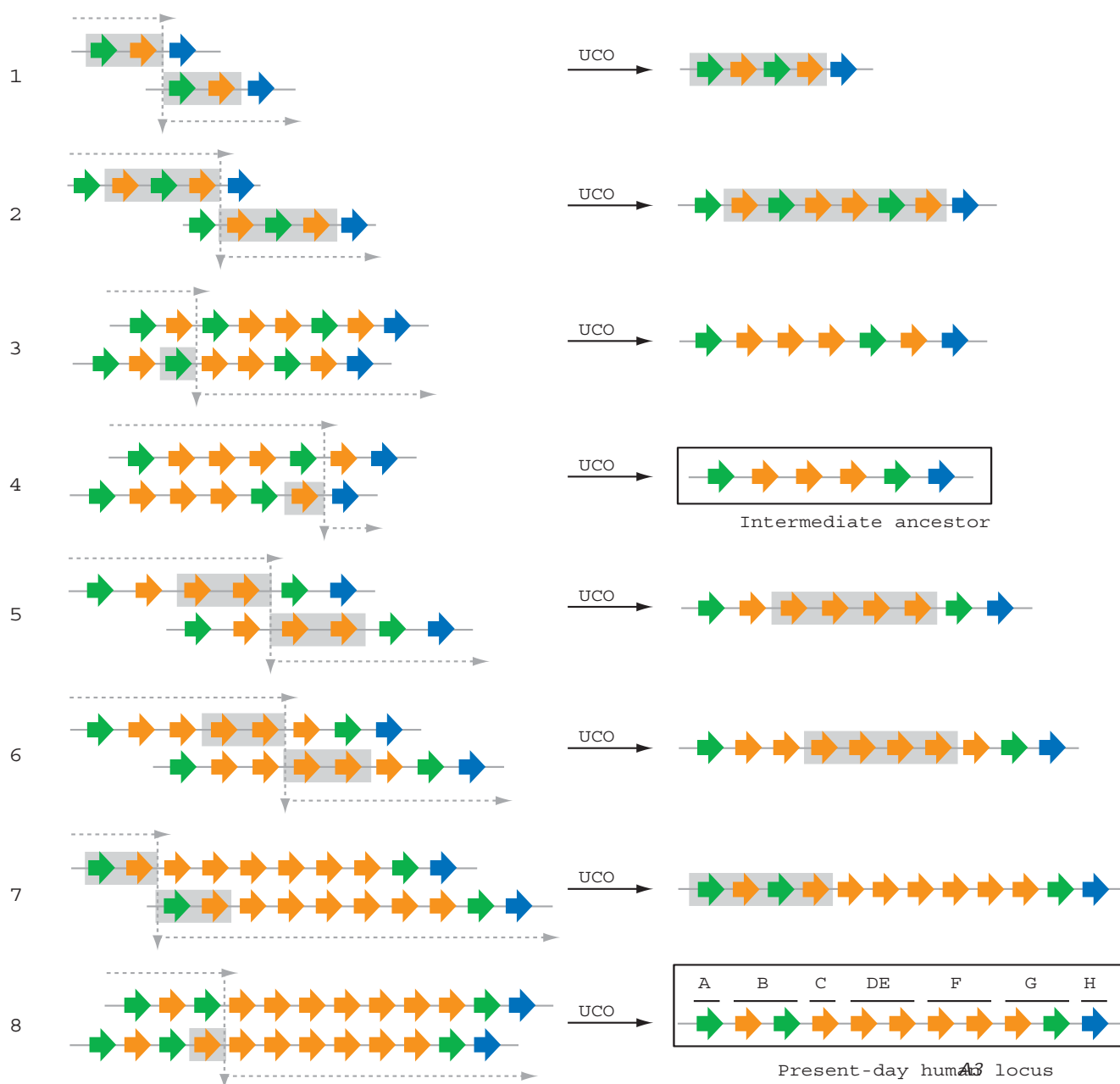
**Figure 6**

**The distribution of A3 Z domains in mammals.** The common ancestor of the indicated placental mammals was inferred to have a Z1-Z2-Z3 A3 gene repertoire. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. A question mark specifies the original *AID*-like ancestor. An asterisk indicates the likely period in which the bulk of the primate A3 gene expansion occurred (see main text, Figure 7 and Additional File 7). Some humans are A3B deficient (minor allele; [75]). The boxed A3 Z domain repertoires constitute the minimal set inferred from incomplete genomic sequences and EST data (Additional File 3).

An 8-event model for human A3 Z domain history is shown in Figure 7 (see Additional File 7 for an alternative representation). This model can be appreciated by considering the present-day human locus and then working backward in time using highly similar local sequences within the A3 locus, which provide 'footprints' for recent recombination events. First, full-length A3A and the Z1 domain of A3B are 97% identical, and they are flanked by nearly homologous ~5.5 kb regions (*i.e.*, direct repeats of 95% identity). These footprints strongly suggested that a recent duplication of two consecutive ancestral domains (Z1-Z2) gave rise to present-day A3B (event 7). Second, we inferred that this recent duplication resulted in a vestigial Z2 domain upstream of A3C, which was subsequently deleted prior to the divergence of human and chimpanzee lineages (event 8). Such a deletion event was supported by the fact that ~3 kb regions of 92% identical DNA reside upstream of the present-day A3B and A3C Z2

domains (these repeats lack similarity to other DNA within the locus). Third, a 92% similarity between two regions (~10 kb) encompassing the A3DE and A3F genes suggested they originated from a recent duplication. Moreover, a similar level of identity was found between two other regions (~10 kb) encompassing the Z2 domains of A3F and A3G. This strongly supported a common ancestral origin for the N-terminal domains of the A3DE, A3F and A3G genes (events 5 and 6). The likelihood of these four relatively recent events suggested that the ancestral locus configuration prior to event 5 [Z1-(Z2)<sub>3</sub>-Z1-Z3] was a key intermediate in the evolution of the primate A3 locus (event 4 product in Figure 7).

Unequal crossing-over events prior to the ancestral intermediate were harder to infer because the footprints have been erased by sequence divergence. We therefore developed an algorithm to compute the minimal series of

**Figure 7**

**An 8-event model for the duplication and deletion history of the human A3 repertoire.** Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. Five duplication and three deletion events were predicted to transform the ancestral locus into the present-day human A3 repertoire. The first event was predicted to occur between two copies of the ancestral Z1-Z2-Z3 locus. The Z domain(s) affected by each unequal crossing-over (UCO) event is shaded gray. The crossing-over points are indicated by a dashed line arrows, and the resulting Z domain configurations are shown (we assumed that new configurations achieved homozygosity prior to being involved in a subsequent UCO). Although deletion events 3 and 4 are illustrated as interchromosomal UCOS, they could have also been caused by intrachromosomal events. Event 4 is depicted before an inferred 'intermediate ancestor' common to nearly all of our models and therefore considered parsimonious, but this event could have occurred any time after event 2. The underlying phylogeny for this model is identical to that shown in Figure 1, except the N-terminal domain of human A3B diverged prior to the point at which the N-terminal domains of human A3F/A3DE and A3G split. An alternative depiction of this model is shown in Additional File 7 and details can be found in the main text and Methods.

duplication and deletion events that could have generated this intermediate locus from the Z1-Z2-Z3 ancestor. Three minimal scenarios were found and each involved 4 events. However, when phylogenetic data were considered, only one scenario was plausible and it involved a 2-domain duplication, a 3-domain duplication and two single domain deletions (respectively, events 1 to 4 in Figure 7 & Additional File 7). Thus, together with the events detailed above, we inferred that the current human A3 repertoire is the product of 8 recombination events – 5 duplications and 3 deletions.

Theoretically, models with as few as 5 events are possible if the likely intermediate locus configuration is ignored. However, these models are also untenable as they clash with phylogenetic and local sequence alignment data. It should be noted that 8 events represent only a lower bound to explain the evolution of the A3 human locus. Scenarios involving more than 8 events could also lead to the same domain organization, and some events may have left no observable trace in the human lineage. Thus, this lower bound could increase when the complete A3 locus sequence of more mammals, and especially more primates, comes available. Finally, it is worth emphasizing that most (if not all) of the 8 recombination events modeled here happened in the 65 to 85 million year period between the points when the rodent and Old World monkey (*e.g.*, rhesus macaque) lineages split from the phylogenetic branch that led to humans (the time frame indicated by the asterisk in Figure 6).

## Discussion

The present studies were initiated to gain a better understanding of the full A3 repertoire of three artiodactyl lineages – cattle, pigs and sheep – and to achieve insights into the mechanism and timing of the A3 gene expansion in mammals. We demonstrated that sheep and cattle have three A3 genes, A3Z1 A3Z2 and A3Z3. However, the latter two genes and their counterpart in pigs have the unique ability to produce a double-domain protein A3Z2-Z3, in addition to single-domain polypeptides. Thus, the A3 proteome of these species is more formidable than gene number alone would indicate. Our studies also help highlight the important point that, although A3 proteins consist of either one or two conserved Z domains, each of these domains can function and evolve independently.

Prior to the present studies, it was clear that most (if not all) placental mammals had Z2- and Z3-type A3 domains (*e.g.*, human, mouse, cat, pig, sheep and cow [35,36,42,54,72]). It was far less clear how broadly the Z1 domain distributed. Here, we presented two critical lines of evidence strongly indicating that the Z1 distribution is equally broad and, importantly, that the common ancestor of placental mammals had a Z1-Z2-Z3 A3 gene reper-

toire, similar to that of present-day sheep and cattle. First, the sheep and cattle A3 genomic sequences demonstrated the presence of a Z1-type A3 gene outside of the primate phylogenetic branches (Figure 6). Second, our pan-species Z1 PCR data, public EST data and draft genomic sequences from horses and dogs combined to show that a A3Z1 gene exists in other parts of the artiodactyl-containing phylogenetic branch set. These data supported a model in which the common ancestor of the primate- and the artiodactyl-containing mammalian super-orders, Euarchotheria and Laurasiatheria, respectively, had a A3Z1 gene and precisely one of each of the three conserved Z domain types (*i.e.*, a divergent model for A3 gene evolution, as opposed to one in which A3Z1 genes evolved independently in several limbs of the mammalian tree). We have therefore established a critical foundation for understanding the function(s) and evolutionary history of the A3 repertoire of any other placental mammal.

It is noteworthy that our pan-species Z1 PCR analyses failed to generate product from opossum genomic DNA and that the recently released opossum and platypus genomic sequences lack A3 genes (Figure 2C; [73,74]). This is unlikely to be a gap in the DNA sequence assemblies because, like non-mammalian vertebrates, DNA and protein searches clearly revealed the A3-flanking genes CBX6 and CBX7 in both animals (LaRue & Harris, unpublished data). Thus, unfortunately, these two interesting non-placental mammals are unlikely to provide significant insights into the earliest stages of A3 gene evolution (*i.e.*, pre-dating the Z1-Z2-Z3 ancestor described here). Perhaps data from the other two placental mammal super-orders, Afrotheria and Xenarthra (*e.g.*, represented by animals such as aardvarks and anteaters, respectively), will help shed light on earlier stages of A3 gene evolution, when presumably an AID-like gene transposed between CBX6 and CBX7 and duplicated to give rise to the ancestral Z1-Z2-Z3 locus. Nevertheless, because all current data indicate that the A3 genes are specific to placental mammals, we hypothesize that a unique role of these genes may relate to the placenta itself, where the A3 proteins may function to help protect the developing fetus from potentially harmful retrotransposition events and/or retroviral infections.

A growing body of evidence indicates that the sole function of the A3 genes of mammals is to provide an innate immune defense to retrovirus and retrotransposon mobilization. This is supported by the fact that the single A3 gene of mice is dispensable and that many of the mammalian A3 genes show evidence for a strong diversifying selection ([10,41,65,66] and this study, Table 1). Although the reason(s) are presently unknown, a large A3 repertoire is clearly more important for some mammals than it is for others. Humans, chimpanzees and rhesus

macaques have 11 Z domains, approximately 3- to 4-fold more than any other known non-primate mammal (Figure 6). Indeed, our studies indicated that the ancestors of humans and chimpanzees experienced at least eight Z domain recombination events, which is more than the total combined number of events for other known mammals. Therefore, despite the fact that the artiodactyl A3 genes show evidence for positive selection, their relative stability in copy number suggests that a considerable disadvantage – such as the potential to mutate genomic DNA – may outweigh the innate benefit of having numerous A3s to combat potentially invasive retroelements. This possibility may very well relate directly to an emerging trend in mammals, which is the frequent loss of a A3Z1 gene which encodes a protein that can penetrate the nuclear compartment (*e.g.*, Figure 5). An A3Z1 deletion was shown here for pigs, inferred here for cats and mice/rats, and demonstrated recently for some human populations (Figure 6 and [75]).

Finally, a major question is what selective pressure(s) drove the A3 expansion from an ancestral Z1-Z2-Z3 repertoire to the present day human Z1-Z2-Z1-(Z2)<sub>6</sub>-Z1-Z3 repertoire? We propose that large-scale events such as gene expansions were selected by extremely pathogenic or lethal retroviral epidemics, because rare expansions would have been easily lost amongst a population of non-expanded alleles. A powerful selective pressure such as a lethal epidemic has the potential to produce a population bottleneck such that mostly (or only) pathogen-resistant individuals would survive (*i.e.*, those with the appropriate disease-resistant A3 repertoire). Such powerful selective pressures would have the potential to promote and perhaps even cause speciation events. We further predict that such events may be marked by changes in A3 Z domain copy number. It is therefore quite plausible that at least some of the eight recombination events required to transform the ancestral Z1-Z2-Z3 repertoire into the present day human Z1-Z2-Z1-(Z2)<sub>6</sub>-Z1-Z3 repertoire may have protected our human ancestors from ancient retroviral infections and thereby facilitated the evolution of primates (a process that we have termed primatification).

## Conclusion

The A3 locus of sheep and cattle consists of three genes, A3Z1, A3Z2 and A3Z3, and the potential to encode four functional proteins, three directly and one (A3Z2-Z3) by read-through transcription and alternative splicing. The A3 locus of pigs experienced a deletion and therefore lacks A3Z1. The artiodactyl A3 repertoire demonstrates a unique modularity centered upon the conserved zinc-coordinating motifs. DNA deaminase activity data and subcellular localization studies suggest that this modularity may also correspond to a broader functionality. All of

the data combined to indicate that the common ancestor of artiodactyls and humans possessed a sheep/cattle-like A3 gene set, with the organization and capacity to evolve into the present day repertoires. The remarkable A3 gene expansion in the primate lineage – from the three ancestral genes (A3Z1-A3Z2-A3Z3) to the present-day eleven Z-domain human repertoire – was predicted to require a minimum of eight recombination events, most of which may have been required to thwart an ancient retroviral infections.

## Methods

### Genomic DNA sequences

A combination of array hybridization, A3-, CBX6- and CBX7-specific PCR was used to identify one A3-positive BACs for sheep (CHORI-243 clone 268D23; a kind gift from P. de Jong, BACPAC Resources Center, <http://bacpac.chori.org/library.php?id=162>) and two for pigs (RPCI-44 clones 344O17 and 408D3; [76]). *E. coli* were transformed with these BACs, grown to saturation in 50 ml cultures and used for DNA preparations as recommended (Marligen Biosciences). Purified BAC DNA was sheared to an average of approximately 3000 bp (Hydroshear method, Genomic Solutions). Fragment ends were blunted with T4 and Klenow DNA polymerases (NEB) and ligated into pBluescriptSK- (Stratagene) or pSMART-HC (Lucigen). Individual subclones were picked randomly and sequenced (ABI3730; Applied Biosystems). Phrap (P. Green, 1996, <http://www.phrap.org/phredphrap/phrap.html>) and Sequencer 4.8 (Gene Codes Corp.) were used to assemble DNA sequences and they were groomed manually. Sequence coverage for the sheep A3 locus averaged 4.5 sequences and the pig 27 sequences. The genomic sequences were compared using Jdotter software (<http://www.jxxi.com/webstart/app/jdotter-a-java-dot-plot-viewer.jsp>; [77]). Repetitive sequences were identified using RepeatMasker <http://www.repeatmasker.org>.

A3 exons were identified by directly comparing the genomic DNA sequences with cDNA, EST and RACE sequences (below, Additional Files 3 & 8 and [42]). Predicted ISREs were identified and compared using the TransFac and Biobase databases through the softberry NSITE portal <http://www.softberry.com>. The sheep CBX6 exons were identified with the help of GenBank EST sequences [EE808826.1](#), [DY519385.1](#) and [EE822736.1](#). The pig CBX6 exons were also identified in this manner using [BP158234.1](#), [BP997823.1](#) & [BP153834.1](#). The sheep and pig CBX7 exons were identified by homology to the cow gene (below). Other CBX6 and CBX7, sequences, respectively, were [NM\\_014292.3](#) and [NM\\_175709.2](#) (human), [NM\\_001103094](#) and [XM\\_604126](#) (cow), [NM\\_028763.3](#) and [NM\\_144811](#) (mouse) and [NM\\_001016617.2](#) & [NM\\_001005071](#) (frog).

### A3Z1 gene degenerate PCR analyses

Genomic DNA was isolated from the following tissues or cell lines: opossum kidney tissue, mouse NIH-3T3 cells, pig PK-15 cells, peccary brain tissue, cow MDBK cells, sheep FLK cells, horse blood cells (PBMC), African green monkey COS7 cells and human 293T cells (DNeasy, Qiagen). 10ng genomic DNA was used as template for PCR using primers designed to anneal to all known A3Z1 genes: 5'-GCC ATG CRG AGC TSY RCT TCY TGG and 5'-GTC ATD ATK GWR AYT YKG GCC CCA GC-3'. Two PCR rounds were used to achieve the final number of cycles (30 plus 18, 21 or 24 cycles). Amplicons were analyzed by agarose gel electrophoresis, TOPO-cloned (Invitrogen) and subjected to DNA sequencing. In all instances, the expected A3Z1 fragments were recovered (*e.g.*, Z1 of human A3A, A3B and A3G could all be detected in a single reaction). 30 PCR cycles using identical conditions and degenerate primers for the ALDOA gene were used as a positive control (5'-CGC TGT GCC CAG TAY AAG AAG GAY GG-3' and 5'-CTG CTG GCA RAT RCT GGC YTA).

### Identifying expressed mRNAs by RACE

RNA was extracted from fresh pig (*Sus scrofa* Landrace/Yorkshire cross), sheep (*Ovis aries* Hampshire) and cattle (*Bos taurus* Hereford) PBMCs using the QIAamp RNA Blood mini kit (Qiagen). 5' and 3' RACE was performed using reagents from the FirstChoice RLM-RACE kit (Ambion). The protocol was modified slightly by using SAP (Roche) instead of CIP to remove 5'-phosphates. A3 cDNA 5' and 3' ends were amplified using Phusion high-fidelity polymerase (NEB), purified and TOPO-cloned (Invitrogen). All A3-specific primers used in conjunction with the 5' and 3' RACE primers are listed in Additional File 8.

### A3 expression plasmids

The pTrc99A-based *E. coli* expression plasmids for sheep, cattle and pig A3Z2-Z3 and for human A3C and A3H were reported previously [42,56]. Other pTrc99A-based constructs were made by ligating KpnI- and SalI-digested PCR fragments into a similarly cut vector. Cow A3Z2 and A3Z3 were amplified from PBMC cDNA (above) using primers 5'-NNN NGA GCT CAG GTA CCA CCA TGC AAC CAG CCT ACC GAG GC & 5'-NNN NGT CGA CTC ACC CGA GAA TGT CCT C and 5'-NNN NGA GCT CAG GTA CCA CCA TGA CCG AGG GCT GGG C & 5'-NNN NGT CGA CCT AAA TTG GGG CCG TTA GGA T, respectively. Pig A3Z2 was amplified from the USMARC1 cDNA library [78] using primers 5'-NNN NGA GCT CAG GTA CCA CCA TGG ATC CTC AGC GCC TGA GAC and 5'-NNN NGT CGA CTC AGC GGT AAC AAA TCC.

Cow A3Z1 was a special case (see main text). It was amplified from PBMC cDNA (above) using primers 5'-NNN NGA GCT CAG GTA CCA C CA TGG ACG AAT ATA CCT

TCA CT and 5'-NNN NGT CGA CGT TTT GCT GAG TCT TGA G and TOPO-cloned into pCR-BLUNT-II-TOPO (Invitrogen). As a control, human A3A was amplified using 5'-NNN NGA GCT CGG TAC CAC CAT GGA AGC CAG CCC AGC and 5'-NNN NGT CGA CCC CAT CCT TCA GTT TCC CTG ATT CTG GAG and TOPO-cloned. Catalytic mutant derivatives of the cow A3Z1 and human A3A plasmids were constructed by site-directed mutagenesis (Stratagene) using oligonucleotides 5'-CCT GCC ATG CAG CGC TCT ACT TCC TG & 5'-CAG GAA GTA GAG CGC TGC ATG GCA GG and 5'-GGC CGC CAT GCG GCG CTG CGT TCT TG & 5'-CAA GAA GCG CAG CGC CGC ATG GCG GCC, respectively.

The artiodactyl A3 proteins were expressed in Hela cells as N-terminal fusions to eGFP (pEGFP-N3; Clontech). Cow and pig A3Z2-Z3-eGFP and the human A3A-, A3C-, A3F- and A3H-eGFP constructs were reported previously [42,79]. Cow and pig A3Z2-eGFP plasmids were made by cloning SacI/SalI-digested PCR products generated using primers 5'-NNN NGA GCT CAG GTA CCA CCA TGC AAC CAG CCT ACC GAG GC & 5'-NNN NGT CGA CCC CGA GAA TGT CCT CAA G and 5'-NNN NGA GCT CAG GTA CCA CCA TGG ATC CTC AGC GCC TGA GAC & 5'-NNN NGT CGA CCC ACC TGG CGT GAG CAC C, respectively. Cow and pig A3Z3-eGFP plasmids were made similarly using primers 5'-NNN NGA GCT CAG GTA CCA CCA TGA CCG AGG GCT GGG C & 5'-NNN GTC GAC TCC AAT TGG GGC CGT TAG GAT and 5'-NNN NGA GCT CAG GTA CCA CCA TGA CCG AGG GCT GGG CT & 5'-NNN GTC GAC TCC TCT CGA GTC ACT TCT TGA, respectively.

Due to the toxicity of cow A3Z1 in *E. coli*, an A3Z1::intron-eGFP plasmid was made by overlapping PCR to join 3 separate fragments: A3Z1 exons 1 and 2 (primers 5'-NNN NGA GCT CAG GTA CCA C CA TGG ACG AAT ATA CCT TCA CT and 5'-CCT GGA CTC ACC TTG TTG CGC), an L1-derived intron ([80]; primers 5'-GTG AGT CCA GGA GAT GTT TCA and 5'-CTG TTG AGA TGA AAG GAG ACA) and A3Z1 exons 3-5 (primers 5'-CAT CTC AAC AGG GTT TGG ATC A and 5'-NNN NGT CGA CGT TTT GCT GAG TCT TGA G). The resulting PCR amplicon was digested with EcoRI and SalI and then ligated into a similarly cut pEGFP-N3 (Clontech).

### Rif<sup>R</sup> DNA deamination assays

Cytosine deaminase activity of the artiodactyl A3 protein variants was measured by quantifying the accumulation of Rif<sup>R</sup> mutants in *ung*-deficient *E. coli* (*e.g.*, [42,56]). All A3 proteins were expressed from pTrc99A (AP Biotech), with the exception of cow A3Z1 and human A3A, which were expressed using pCR-BLUNT-II-TOPO (Invitrogen). Experiments were done a minimum of three times, in the presence or absence of IPTG as indicated.

### Fluorescence microscopy

To observe subcellular localization of A3 proteins, 5000 HeLa cells were incubated for 24 h in Labtek chambered coverglasses (Nunc), transfected with 200 ng of the pEGFP-N3 based constructs and, after an additional 24 h visualized on a Zeiss Axiovert 200 microscope at 400× magnification. HsA3F, HsA3C, HsA3H, HsA3A, BtA3Z2-Z3 and SsA3Z2-Z3 fusion constructs were previously reported [19,42,79].

### Phylogenies and positive selection calculations

Z domain exons were used for all phylogenetic, positive selection and modelling studies. GARD showed no evidence for recombination breakpoints within the Z domain exons [81]. T<sub>coffee</sub> version 5.31 was used for multiple sequence alignments [82]. PAL2NAL software was used to convert amino acid sequences to nucleotides [83]. JalView was used to remove insertions/deletions [84]. The dnaml program within the Phylip software package was used to generate a phylogenetic tree ([85]; an identical tree was obtained with MrBayes version 3.1 [86], except branch lengths differed slightly). Clustal W version 1.83.1 was also used for some individual domain comparisons [87].

Free ratio model positive selection studies were based on a phylogenetic tree generated through Bayesian inference using MrBayes version 3.1 [86]. Each tree was run for 250,000 generations with a burnin of 62,500 and standard default parameters. The PAML codeml program [88] was used to generate dN/dS ratios ( $\omega$  values) for phylogenetic tree branches.  $\omega$  values from the free ratio model using the F3 × 4 algorithm are shown in Additional File 6 (values from the F1 × 61 algorithm were similar and therefore not shown).

Positive selection was also evaluated in specific phylogenetic lineages using the NsSites model in the PAML codeml program (Table 1). Individual Z domain phylogenetic trees were generated as described above and used in these analyses. Z2 and Z3 comparisons were done for sheep, cow, pig, peccary and horse sequences, and Z1 comparisons for sheep, cow, horse and dog sequences (non-artiodactyl sequences were added for statistical significance; GenBank accession numbers are in Additional File 3). Models for neutral selection (M1 and M7) were compared to those for positive selection (M2 and M8). Likelihood ratio tests were performed to compare the null and positive selection scenarios.

### A3 gene expansion modelling

The aim was to infer the most likely histories of duplications and deletions that gave rise to the human A3 locus. Instead of considering each gene as an individual element,

we subdivided it into its N-terminal and C-terminal Z-domains. Hence, the present-day human locus configuration was represented as follows: Z1-Z2-Z1-(Z2)<sub>6</sub>-Z1-Z3. The considered duplications are 'multiple tandem duplications' resulting from unequal crossing over [70]. In other words, a single duplication event can copy an arbitrary number of consecutive Z-domains, and place them in the same order next to the original ones. Similarly, an unequal crossing-over can remove an arbitrary number of adjacent domains and cause deletions.

Various algorithms have been proposed to infer evolutionary histories of tandemly arrayed gene families [71,89-91], but none of them involve both multiple tandem duplications and deletions. Consequently, we developed a brute force algorithm to enumerate all possible evolutionary scenarios involving a minimum number of duplications and deletions that can transform a particular locus configuration into another. Such an exhaustive algorithm has an exponential time complexity and it is impractical for analyzing large gene families. However, the limited size of the A3 locus and the classification of the Z domains into three distinct categories made it useful here (e.g., events 1 to 4 in Figure 7 & Additional File 7).

To infer the most recent evolutionary events (events 5 to 8 in Figure 7 & Additional File 7), we performed an analysis of the self-similarities within the human A3 locus. The DNA sequence (hg18, chr22:37682569-37830946) with identified interspersed repeats was downloaded from the RepeatMasker web site <http://www.repeatmasker.org>. A dot plot of this sequence with itself was obtained using Gepard [92] to identify pairs of regions with very high similarities. The three most significant were extracted and further aligned using Blastz [93] with default parameter to obtain the percentage of identity. These regions were used to infer and model the most recent evolutionary events, as described in the main text.

### Data deposition

The GenBank accession number for the sheep A3 genomic sequence is [FJ042940](#). The GenBank accession numbers for the two pig A3 genomic sequences are [FJ042938](#) and [FJ042939](#). All A3 cDNA and EST sequences have also been deposited (see Additional File 3 for a full list of GenBank accession numbers).

### Abbreviations

A.A.: amino acid; A3: APOBEC3; A3A: APOBEC3B; A3B: APOBEC3C; A3C: APOBEC3C; A3DE: APOBEC3DE; A3F: APOBEC3F; A3G: APOBEC3G; A3H: APOBEC3H; GFP: Green Fluorescent Protein; PAML: Phylogenetic Analysis by Maximum Likelihood; MYA: Millions of Years Ago; Z: Zinc-coordinating motif

## Authors' contributions

RSL and RSH designed the studies, performed experiments, analyzed data and wrote the manuscript. SRJ and VA helped analyze the artiodactyl A3 genes and proteins, TPLS provided library samples and generated genomic DNA sequences, IH contributed cattle A3 gene sequences and functional data, KATS assisted with phylogenetic and computational studies, and ML, DB and NE generated the model for A3 evolution. All authors contributed to editing the manuscript.

## Additional material

### Additional file 1

**APOBEC3 Z domain conservation.** Web LOGO profiles depicting amino acid conservation within each mammalian Z domain. The multiple sequence alignments used to generate the phylogenetic tree in Figure 1 were used to create consensus profiles for each of the indicated Z domains using Web LOGO [94]. Arrowheads below the amino acid profiles indicate residues that define each Z type (see the main text for additional details).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S1.ps>]

### Additional file 2

**APOBEC3 genomic locus comparisons.** Dotplot alignments of (A) the sheep and pig and (B) the sheep and cow A3 genomic sequences.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S2.eps>]

### Additional file 3

**Mammalian A3 and AID sequences.** A table summarizing all DNA sequences used in this study, including GenBank accession numbers.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S3.eps>]

### Additional file 4

**APOBEC3 promoter element conservation.** Predicted interferon-stimulating response elements (ISRE) in the promoter regions of the indicated A3 genes and known interferon-inducing genes ISG54 and ISG15. The ISRE sequences are shown relative to the translation initiation codon ATG. Identities to human sequences are shaded gray.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S4.eps>]

### Additional file 5

**E. coli-based DNA cytosine deaminase activity data.** DNA cytosine deaminase activity of the pig A3Z2-Z3 and A3Z2 proteins in E. coli. Conditions and labels are identical to those used in Figure 4, except 10 independent cultures were grown under IPTG-induced conditions and analyzed.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S5.eps>]

### Additional file 6

**Evidence for positive selection in APOBEC3 gene evolution.** Phylogenetic trees showing relative relationships and  $\omega$  values for the indicated (A) Z1 domains, (B) Z2 domains, (C) Z3 domains and (D) the Z domain of AID. The phylogenetic trees were determined using MrBayes, and the  $\omega$  values were calculated using the PAML free ratio model.  $\omega$  values are shown in red adjacent to (or where space is non-permitting, to right of) each phylogenetic branch. Asterisks denote branches where the  $\omega$  value was infinity (i.e., dS was zero). The units for the scale bars are nucleotide changes per codon. The dotted line in panel (C) was used to provide more space to depict the human and non-human primate Z3 tree branches. See the main text and Methods for additional details.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S6.eps>]

### Additional file 7

**Proposed APOBEC3 gene diversification events during primate evolution.** An alternative representation of the 8-event model for the duplication and deletion history of the human A3 repertoire. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. The Z domain(s) involved in each event are shaded gray. Dark black and red lines mark duplications (one color for the original segment and one color for the duplicated segment), crosses designate deletions and light gray lines indicate no change. See the main text, Figure 7 and Methods for details.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S7.doc>]

### Additional file 8

**Primers used to identify expressed APOBEC3 transcripts from cow, sheep and pig PBMCs.** A table summarizing the oligonucleotide primers used in this study.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S8.doc>]

## Acknowledgements

We thank L. Beach, H. Malik, M. Murtaugh and M. Stenglein for valuable feedback. We thank K. Tennill and R. Godtel for assistance with BAC DNA sequencing, D. Shiroma and S. Fahrenkrug for help identifying pig BAC clones, M. Stenglein for several expression plasmids, M. Titus for use of her microscope, L. Hartman for pig and sheep samples, C. Knutson for cow samples, J. Zimmerman and R. Molina for peccary blood, P. Krauseman for peccary brain tissue and M. Ruen and O. Holland for opossum samples. R. LaRue is a member of the University of Minnesota CMB Graduate Program. S. Jónsson was the 2004–2005 Val Bjornson Icelandic Exchange Scholarship recipient. M. Lajoie was supported by a Canadian Institutes of Health Research studentship. D. Bertrand and N. El-Mabrouk were supported by grants from the Fonds Québécois de la Recherche sur la Nature et les Technologies and the Natural Sciences and Engineering Research Council of Canada. R. Harris was supported in part by a Searle Scholarship and a University of Minnesota McKnight Land Grant Assistant Professorship. This work was also supported by NIH grant AI064046. The University of Minnesota Advanced Genetic Analysis Facility assisted with DNA sequencing.

## References

1. Bieniasz PD: **Intrinsic immunity: a front-line defense against viral attack.** *Nat Immunol* 2004, **5**(11):1109–1115.

2. Chiu YL, Greene WC: **The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements.** *Annu Rev Immunol* 2008, **26**:317-353.
3. Cullen BR: **Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors.** *J Virol* 2006, **80**(3):1067-1076.
4. Goff SP: **Retrovirus restriction factors.** *Mol Cell* 2004, **16**(6):849-859.
5. Malim MH, Emerman M: **HIV-1 accessory proteins – ensuring viral survival in a hostile environment.** *Cell Host Microbe* 2008, **3**(6):388-398.
6. Doeble BP, Bogerd HP, Wiegand HL, Jouvenet N, Bieniasz PD, Hunter E, Cullen BR: **The betaretrovirus Mason-Pfizer monkey virus selectively excludes simian APOBEC3G from virion particles.** *J Virol* 2006, **80**(24):12102-12108.
7. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH: **DNA deamination mediates innate immunity to retroviral infection.** *Cell* 2003, **113**(6):803-809.
8. Jost S, Turelli P, Mangeat B, Protzer U, Trono D: **Induction of anti-viral cytidine deaminases does not explain the inhibition of hepatitis B virus replication by interferons.** *J Virol* 2007, **81**(19):10588-10596.
9. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D: **Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts.** *Nature* 2003, **424**(6944):99-103.
10. Okeoma CM, Lovsin N, Peterlin BM, Ross SR: **APOBEC3 inhibits mouse mammary tumour virus replication in vivo.** *Nature* 2007, **445**(7130):927-930.
11. Suspene R, Guetard D, Henry M, Sommer P, Wain-Hobson S, Vartanian JP: **Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo.** *Proc Natl Acad Sci USA* 2005, **102**(23):8321-8326.
12. Turelli P, Mangeat B, Jost S, Vianin S, Trono D: **Inhibition of hepatitis B virus replication by APOBEC3G.** *Science* 2004, **303**(5665):1829.
13. Wiegand HL, Cullen BR: **Inhibition of alpharetrovirus replication by a range of human APOBEC3 proteins.** *J Virol* 2007, **81**(24):13694-13699.
14. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L: **The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA.** *Nature* 2003, **424**(6944):94-98.
15. Armitage AE, Katourakis A, de Oliveira T, Welch JJ, Belshaw R, Bishop KN, Kramer B, McMichael AJ, Rambaut A, Iversen AK: **Conserved footprints of APOBEC3G on hypermutated HIV-1 and HERV-K(HML2) sequences.** *J Virol* 2008, **82**(17):8743-8761.
16. Dutko JA, Schafer A, Kenny AE, Cullen BR, Curcio MJ: **Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases.** *Curr Biol* 2005, **15**(7):661-666.
17. Esnault C, Heidmann O, Delebecque F, Dewannieux M, Ribet D, Hance AJ, Heidmann T, Schwartz O: **APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses.** *Nature* 2005, **433**(7024):430-433.
18. Jern P, Stoye JP, Coffin JM: **Role of APOBEC3 in genetic diversity among endogenous murine leukemia viruses.** *PLoS Genet* 2007, **3**(10):2014-2022.
19. Jónsson SR, LaRue RS, Stenglein MD, Fahrenkrug SC, Andrésdóttir V, Harris RS: **The restriction of zoonotic PERV transmission by human APOBEC3G.** *PLoS ONE* 2007, **2**(9):e893.
20. Kaiser SM, Malik HS, Emerman M: **Restriction of an extinct retrovirus by the human TRIM5alpha antiviral protein.** *Science* 2007, **316**(5832):1756-1758.
21. Lee YN, Bieniasz PD: **Reconstitution of an infectious human endogenous retrovirus.** *PLoS Pathog* 2007, **3**(1):e10.
22. Lee YN, Malim MH, Bieniasz PD: **Hypermutation of an ancient human retrovirus by APOBEC3G.** *J Virol* 2008, **17**:8762-8770.
23. Schumacher AJ, Nissley DV, Harris RS: **APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast.** *Proc Natl Acad Sci USA* 2005, **102**(28):9854-9859.
24. Chen H, Lilley CE, Yu Q, Lee DV, Chou J, Narvaiza I, Landau NR, Weitzman MD: **APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons.** *Curr Biol* 2006, **16**(5):480-485.
25. Kinomoto M, Kanno T, Shimura M, Ishizaka Y, Kojima A, Kurata T, Sata T, Tokunaga K: **All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition.** *Nucleic Acids Res* 2007, **35**(9):2955-2964.
26. Turelli P, Vianin S, Trono D: **The innate antiretroviral factor APOBEC3G does not affect human LINE-1 retrotransposition in a cell culture assay.** *J Biol Chem* 2004, **279**(42):43371-43373.
27. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, Flory E, Schumann GG, Munk C: **APOBEC3 proteins inhibit human LINE-1 retrotransposition.** *J Biol Chem* 2006, **281**(31):22161-22172.
28. Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, Cullen BR: **Cellular inhibitors of long interspersed element 1 and Alu retrotransposition.** *Proc Natl Acad Sci USA* 2006, **103**(23):8780-8785.
29. Niewiadomska AM, Tian C, Tan L, Wang T, Thi Nguyen Sarkis P, Yu XF: **Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with HMM complex formation or P-body association.** *J Virol* 2007, **17**:9577-9583.
30. Stenglein MD, Harris RS: **APOBEC3B and APOBEC3F inhibit LI retrotransposition by a DNA deamination-independent mechanism.** *J Biol Chem* 2006, **281**(25):16837-16841.
31. Abudu A, Takaori-Kondo A, Izumi T, Shirakawa K, Kobayashi M, Sasada A, Fukunaga K, Uchiyama T: **Murine retrovirus escapes from murine APOBEC3 via two distinct novel mechanisms.** *Curr Biol* 2006, **16**(15):1565-1570.
32. Derse D, Hill SA, Princlair G, Lloyd P, Heidecker G: **Resistance of human T cell leukemia virus type 1 to APOBEC3G restriction is mediated by elements in nucleocapsid.** *Proc Natl Acad Sci USA* 2007, **104**(8):2915-2920.
33. Russell RA, Wiegand HL, Moore MD, Schafer A, McClure MO, Cullen BR: **Foamy virus Bet proteins function as novel inhibitors of the APOBEC3 family of innate antiretroviral defense factors.** *J Virol* 2005, **79**(14):8724-8731.
34. Yu X, Yu Y, Liu B, Luo K, Kong W, Mao P, Yu XF: **Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex.** *Science* 2003, **302**(5647):1056-1060.
35. Conticello SG, Thomas CJ, Petersen-Mahrt S, Neuberger MS: **Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases.** *Mol Biol Evol* 2005, **22**(2):367-377.
36. Harris RS, Liddament MT: **Retroviral restriction by APOBEC proteins.** *Nat Rev Immunol* 2004, **4**(11):868-877.
37. Larue RS, Andrésdóttir V, Blanchard Y, Conticello SG, Derse D, Emerman M, Greene WC, Jónsson SR, Landau NR, Löchelt M, et al.: **Guidelines for naming non-primate APOBEC3 genes and proteins.** *J Virol* 2008 in press.
38. Haché G, Liddament MT, Harris RS: **The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain.** *J Biol Chem* 2005, **280**(12):10920-10924.
39. Navarro F, Bollman B, Chen H, König R, Yu Q, Chiles K, Landau NR: **Complementary function of the two catalytic domains of APOBEC3G.** *Virology* 2005, **333**(2):374-386.
40. Newman EN, Holmes RK, Craig HM, Klein KC, Lingappa JR, Malim MH, Sheehy AM: **Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity.** *Curr Biol* 2005, **15**(2):166-170.
41. OhAinle M, Kerns JA, Malik HS, Emerman M: **Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H.** *J Virol* 2006, **80**(8):3853-3862.
42. Jónsson SR, Haché G, Stenglein MD, Fahrenkrug SC, Andrésdóttir V, Harris RS: **Evolutionarily conserved and non-conserved retrovirus restriction activities of artiodactyl APOBEC3F proteins.** *Nucleic Acids Res* 2006, **34**(19):5683-5694.
43. Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A: **The delayed rise of present-day mammals.** *Nature* 2007, **446**(7135):507-512.
44. Kumar S, Hedges SB: **A molecular timescale for vertebrate evolution.** *Nature* 1998, **392**(6679):917-920.
45. Reza Shariflou M, Moran C: **Conservation within artiodactyls of an AATA interrupt in the IGF-1 microsatellite for 19–35 million years.** *Mol Biol Evol* 2000, **17**(4):665-669.
46. Peng G, Lei KJ, Jin W, Greenwell-Wild T, Wahl SM: **Induction of APOBEC3 family proteins, a defensive maneuver underlying**

- interferon-induced anti-HIV-1 activity. *J Exp Med* 2006, **203**(1):41-46.
47. Tanaka Y, Marusawa H, Seno H, Matsumoto Y, Ueda Y, Kodama Y, Endo Y, Yamauchi J, Matsumoto T, Takaori-Kondo A, et al.: **Antiviral protein APOBEC3G is induced by interferon-alpha stimulation in human hepatocytes.** *Biochem Biophys Res Commun* 2006, **341**(2):314-319.
  48. Abrahams VM, Schaefer TM, Fahey JV, Visintin I, Wright JA, Aldo PB, Romero R, Wira CR, Mor G: **Expression and secretion of antiviral factors by trophoblast cells following stimulation by the TLR-3 agonist, Poly(I : C).** *Hum Reprod* 2006, **21**(9):2432-2439.
  49. Bonvin M, Achermann F, Greeve I, Stroka D, Keogh A, Inderbitzin D, Candinas D, Sommer P, Wain-Hobson S, Vartanian JP, et al.: **Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication.** *Hepatology* 2006, **43**(6):1364-1374.
  50. Kessler DS, Levy DE, Darnell JE Jr: **Two interferon-induced nuclear factors bind a single promoter element in interferon-stimulated genes.** *Proc Natl Acad Sci USA* 1988, **85**(22):8521-8525.
  51. Levy DE, Kessler DS, Pine R, Reich N, Darnell JE Jr: **Interferon-induced nuclear factors that bind a shared promoter element correlate with positive and negative transcriptional control.** *Genes Dev* 1988, **2**(4):383-393.
  52. Reich N, Evans B, Levy D, Fahey D, Knight E Jr, Darnell JE Jr: **Interferon-induced transcription of a gene encoding a 15-kDa protein depends on an upstream enhancer element.** *Proc Natl Acad Sci USA* 1987, **84**(18):6394-6398.
  53. Muckenfuss H, Kaiser JK, Krebil E, Battenberg M, Schwer C, Cichutek K, Munk C, Flory E: **Sp1 and Sp3 regulate basal transcription of the human APOBEC3G gene.** *Nucleic Acids Res* 2007, **35**(11):3784-3796.
  54. Munk C, Beck T, Zielonka J, Hotz-Wagenblatt A, Chareza S, Battenberg M, Thielebein J, Cichutek K, Bravo IG, O'Brien SJ, et al.: **Functions, structure, and read-through alternative splicing of feline APOBEC3 genes.** *Genome Biol* 2008, **9**(3):R48.
  55. Dang Y, Wang X, Esselman WJ, Zheng YH: **Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family.** *J Virol* 2006, **80**(21):10522-10533.
  56. Harris RS, Petersen-Mahrt SK, Neuberger MS: **RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators.** *Molecular Cell* 2002, **10**(5):1247-1253.
  57. Liddament MT, Brown WL, Schumacher AJ, Harris RS: **APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo.** *Curr Biol* 2004, **14**(15):1385-1391.
  58. Mariani R, Chen D, Schröfelbauer B, Navarro F, König R, Bollman B, Munk C, Nymark-McMahon H, Landau NR: **Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif.** *Cell* 2003, **114**(1):21-31.
  59. Yu Q, Chen D, König R, Mariani R, Unutmaz D, Landau NR: **APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication.** *J Biol Chem* 2004, **279**(51):53379-53386.
  60. Iwatani Y, Takeuchi H, Strebel K, Levin JG: **Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect.** *J Virol* 2006, **80**(12):5992-6002.
  61. Bogerd HP, Wiegand HL, Doehle BP, Cullen BR: **The intrinsic antiretroviral factor APOBEC3B contains two enzymatically active cytidine deaminase domains.** *Virology* 2007, **364**(2):486-493.
  62. Dang Y, Siew LM, Wang X, Han Y, Lampen R, Zheng YH: **Human cytidine deaminase APOBEC3H restricts HIV-1 replication.** *J Biol Chem* 2008, **283**(17):11606-11614.
  63. OhAinle M, Kerns JA, Li MM, Malik HS, Emerman M: **Antiretroelement activity of APOBEC3H was lost twice in recent human evolution.** *Cell Host Microbe* 2008, **4**(3):249-259.
  64. Chen KM, Martemyanova N, Lu Y, Shindo K, Matsuo H, Harris RS: **Extensive mutagenesis experiments corroborate a structural model for the DNA deaminase domain of APOBEC3G.** *FEBS Lett* 2007, **581**(24):4761-4766.
  65. Sawyer SL, Emerman M, Malik HS: **Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G.** *PLoS Biol* 2004, **2**(9):E275.
  66. Zhang J, Webb DM: **Rapid evolution of primate antiviral enzyme APOBEC3G.** *Hum Mol Genet* 2004, **13**(16):1785-1791.
  67. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, Wilson RK, et al.: **Evolutionary and biomedical insights from the rhesus macaque genome.** *Science* 2007, **316**(5822):222-234.
  68. Virgen CA, Hatzioannou T: **Antiretroviral activity and Vif sensitivity of rhesus macaque APOBEC3 proteins.** *J Virol* 2007, **81**(24):13932-13937.
  69. Purvis A: **A composite estimate of primate phylogeny.** *Philos Trans R Soc Lond B Biol Sci* 1995, **348**(1326):405-421.
  70. Fitch WM: **Phylogenies constrained by the crossover process as illustrated by human hemoglobins and a thirteen-cycle, eleven-amino-acid repeat in human apolipoprotein A-I.** *Genetics* 1977, **86**(3):623-644.
  71. Gascuel O, Bertrand D, Elemento O: **Reconstructing the duplication history of tandemly repeated sequences.** *"Mathematics of Evolution and Phylogeny"* 2007:205-235.
  72. Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, Scott J, Navaratnam N: **An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22.** *Genomics* 2002, **79**(3):285-296.
  73. Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, et al.: **Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences.** *Nature* 2007, **447**(7141):167-177.
  74. Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grutzner F, Belov K, Miller W, Clarke L, Chinwalla AT, et al.: **Genome analysis of the platypus reveals unique signatures of evolution.** *Nature* 2008, **453**(7192):175-183.
  75. Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE: **Population stratification of a common APOBEC gene deletion polymorphism.** *PLoS Genet* 2007, **3**(4):e63.
  76. Fahrenkrug SC, Rohrer GA, Freking BA, Smith TP, Osoegawa K, Shu CL, Catanese JJ, de Jong PJ: **A porcine BAC library with tenfold genome coverage: a resource for physical and genetic map integration.** *Mamm Genome* 2001, **12**(6):472-474.
  77. Brodie R, Smith AJ, Roper RL, Tcherepanov V, Upton C: **Base-By-Base: single nucleotide-level analysis of whole viral genome alignments.** *BMC Bioinformatics* 2004, **5**:96.
  78. Fahrenkrug SC, Smith TP, Freking BA, Cho J, White J, Vallet J, Wise T, Rohrer G, Perteau G, Sultana R, et al.: **Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly.** *Mamm Genome* 2002, **13**(8):475-478.
  79. Stenglein MD, Matsuo H, Harris RS: **Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization.** *J Virol* 2008, **82**(19):9591-9599.
  80. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazanian HH Jr: **High frequency retrotransposition in cultured mammalian cells.** *Cell* 1996, **87**(5):917-927.
  81. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelck CH, Frost SD: **GARD: a genetic algorithm for recombination detection.** *Bioinformatics* 2006, **22**(24):3096-3098.
  82. Notredame C, Higgins DG, Heringa J: **T-Coffee: A novel method for fast and accurate multiple sequence alignment.** *J Mol Biol* 2000, **302**(1):205-217.
  83. Suyama M, Torrents D, Bork P: **PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments.** *Nucleic Acids Res* 2006:W609-612.
  84. Clamp M, Cuff J, Searle SM, Barton GJ: **The Jalview Java alignment editor.** *Bioinformatics* 2004, **20**(3):426-427.
  85. Retief JD: **Phylogenetic analysis using PHYLIP.** *Methods Mol Biol* 2000, **132**:243-258.
  86. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**(12):1572-1574.
  87. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al.: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23**(21):2947-2948.
  88. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 1997, **13**(5):555-556.
  89. Bertrand D, Lajoie M, El-Mabrouk N: **Inferring ancestral gene orders for a family of tandemly arrayed genes.** *J Comp Biol* 2008, **15**(8):1063-1077.
  90. Lajoie M, Bertrand D, El-Mabrouk N, Gascuel O: **Duplication and inversion history of a tandemly repeated genes family.** *J Comput Biol* 2007, **14**(4):462-478.

91. Zhang L, Ma B, Wang L, Xu Y: **Greedy method for inferring tandem duplication history.** *Bioinformatics* 2003, **19(12)**:1497-1504.
92. Krumsiek J, Arnold R, Rattei T: **Gepard: a rapid and sensitive tool for creating dotplots on genome scale.** *Bioinformatics* 2007, **23(8)**:1026-1028.
93. Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W: **Human-mouse alignments with BLASTZ.** *Genome Res* 2003, **13(1)**:103-107.
94. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14(6)**:1188-1190.

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)



## Supplemental figure legends

**Figure S1.** Web LOGO profiles depicting amino acid conservation within each mammalian Z domain. The multiple sequence alignments used to generate the phylogenetic tree in **Figure 1** were used to create consensus profiles for each of the indicated Z domains using Web LOGO [94]. Arrowheads below the amino acid profiles indicate residues that define each Z type (see the main text for additional details).

**Figure S2.** Dotplot alignments of (A) the sheep and pig and (B) the sheep and cow *A3* genomic sequences.

**Figure S3.** Predicted interferon-stimulating response elements (ISRE) in the promoter regions of the indicated *A3* genes and known interferon-inducing genes *ISG54* and *ISG15*. The ISRE sequences are shown relative to the translation initiation codon ATG. Identities to human sequences are shaded gray.

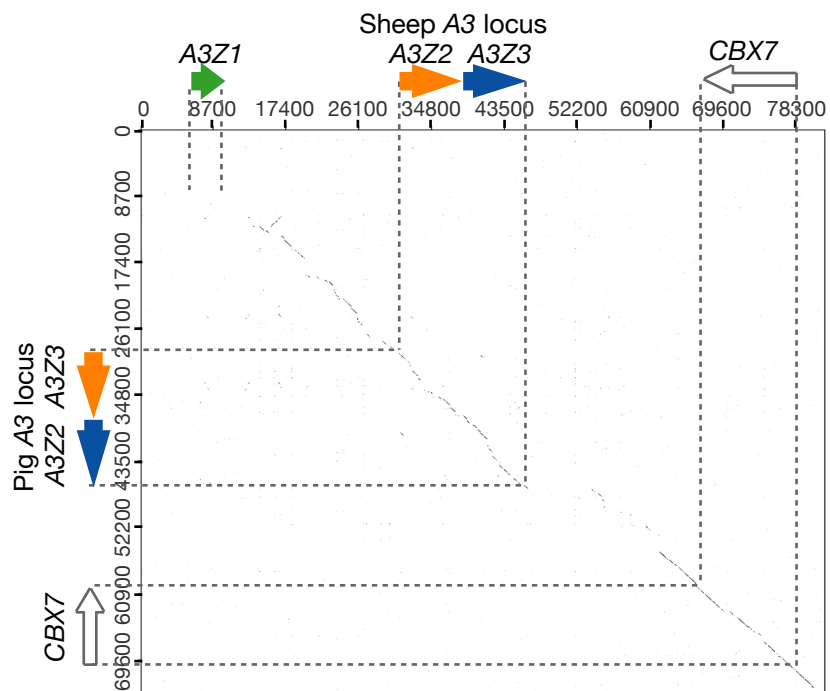
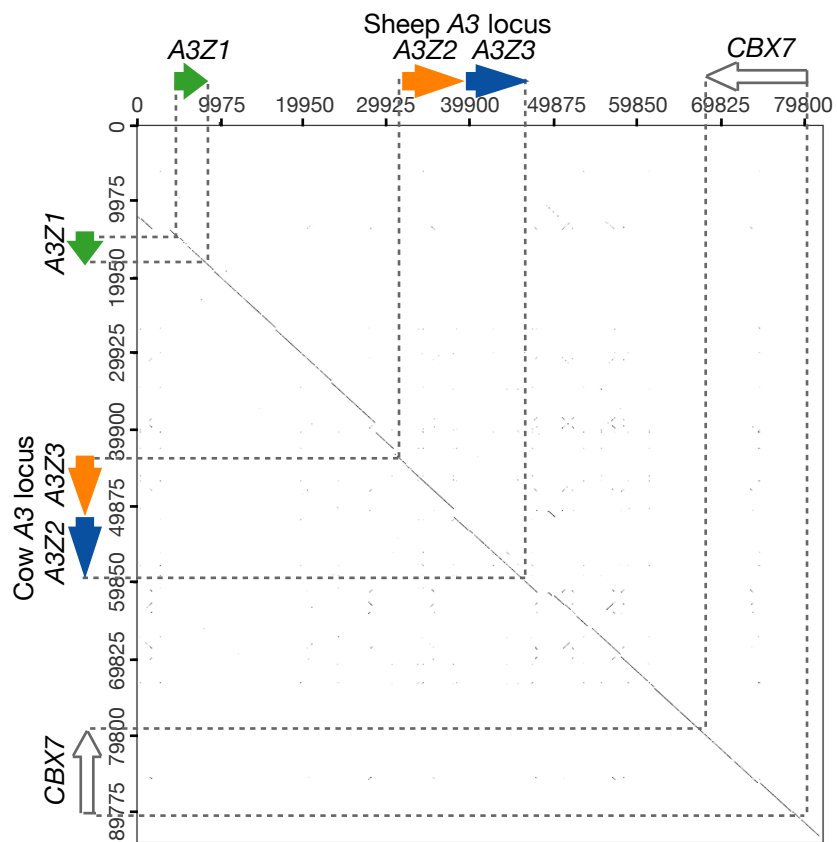
**Figure S4.** DNA cytosine deaminase activity of the pig A3Z2-Z3 and A3Z2 proteins in *E. coli*. Conditions and labels are identical to those used in **Figure 4**, except 10 independent cultures were grown under IPTG-induced conditions and analyzed.

**Figure S5.** Phylogenetic trees showing relative relationships and  $\omega$  values for the indicated (A) Z1 domains, (B) Z2 domains, (C) Z3 domains and (D) the Z domain of *AID*. The phylogenetic trees were determined using MrBayes, and the  $\omega$  values were

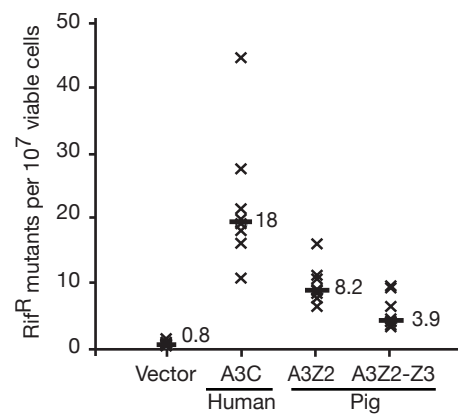
calculated using the PAML free ratio model.  $\omega$  values are shown in red adjacent to (or where space is non-permitting, to right of) each phylogenetic branch. Asterisks denote branches where the  $\omega$  value was infinity (*i.e.*, dS was zero). The units for the scale bars are nucleotide changes per codon. The dotted line in panel (C) was used to provide more space to depict the human and non-human primate Z3 tree branches. See the main text and **Methods** for additional details.

**Figure S6.** An alternative representation of the 8-event model for the duplication and deletion history of the human *A3* repertoire. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. The Z domain(s) involved in each event are shaded gray. Dark black and red lines mark duplications (one color for the original segment and one color for the duplicated segment), crosses designate deletions and light gray lines indicate no change. See the main text, **Figure 7** and **Methods** for details.

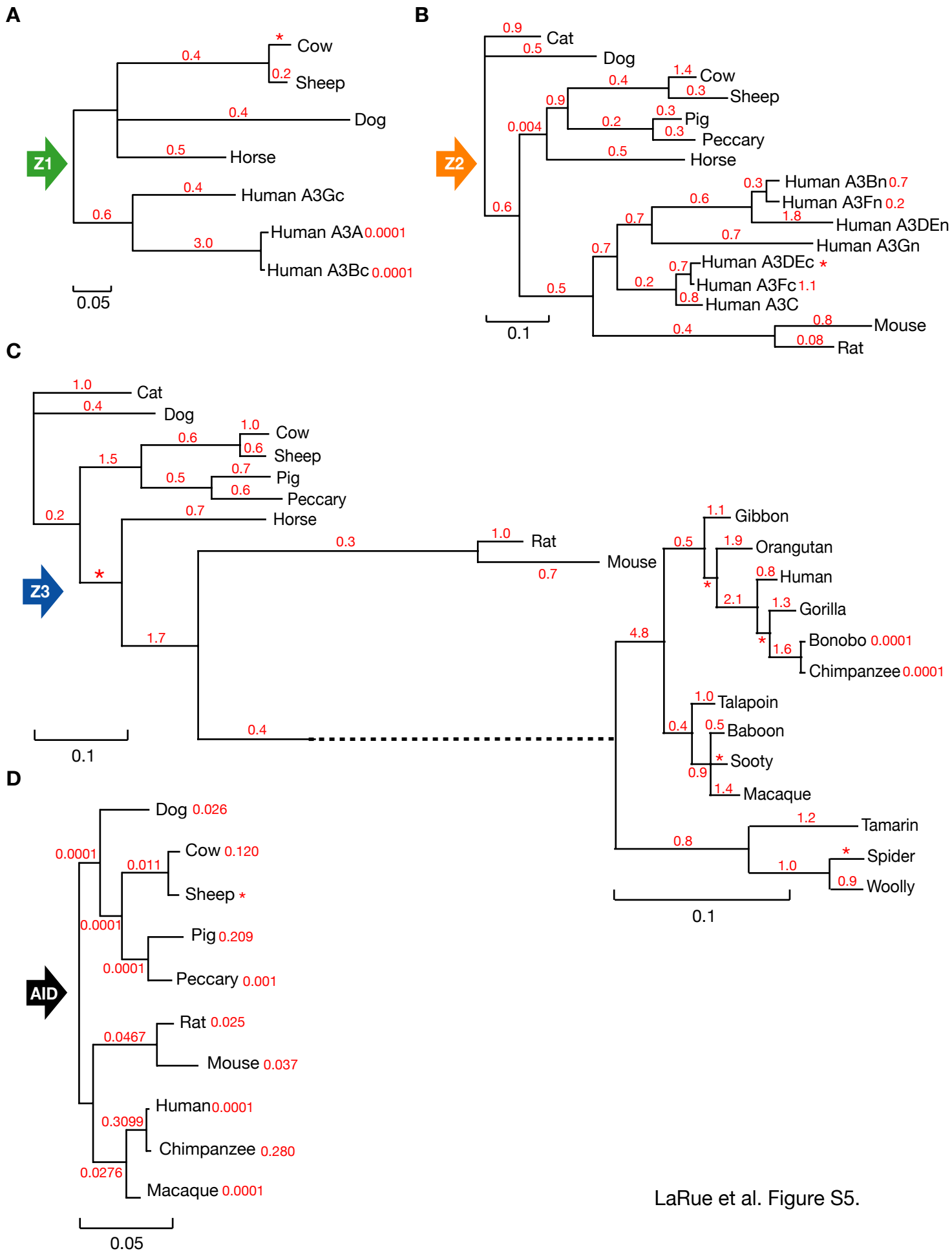


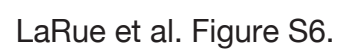
**A****B**

<b>A3 gene</b>	<b>ISRE</b>	<b>Distance from ATG</b>
Sheep A3Z2	TTTACTTTCTCTTTCCCTTT	-311
Cow A3Z2	TTTACTTTCTCTTTCCCTTT	-282
Pig A3Z2	TTTACTTTCTCTTTCCCTTT	-230
Cat A3Z2a	CCTCCTTTCTCTTTCCCTTT	-214
Cat A3Z2b	AGTACTTTCTCTTTCCCTTT	-246
Cat A3Z2c	AGTACTTTCTCTTTCCCTTT	-244
Human A3DE, A3F, A3G	- - GACTTTCTCTTTCCCTTT	-602, -283, -284
Sheep A3Z3	TTTACTTTCTGTTTCCCCTT-	-208
Cow A3Z3	TTTACTTTCTCTTTCCCCTTT	-214
Pig A3Z3	TTTACTTTCTCTTTCCC- TTT	-232
Cat A3Z3	AGTACTTTCTCTTTCTC- TTT	-243
Human A3DE, A3F, A3G	- - GACTTTCTCTTTCCC- TTT	-602, -283, -284
Human ISG54	TCTAGTTTCACTTTCCCTTT	
Human ISG15	- - CAGTTTCGGTTTCCCTTT	



LaRue et al. Figure S4.





**Table S1.** Mammalian A3 and AID sequences

Species	Common name	Protein	Accession number	Journal
<i>Bos taurus</i>	Cow	A3Z1	EU864534	This study
<i>Bos taurus</i>	Cow	A3Z2-Z3	DQ974646	Jónsson <i>et al.</i> , 2006
<i>Bos taurus</i>	Cow	A3Z2	EU864535	This study
<i>Bos taurus</i>	Cow	A3Z3	EU864536	This study
<i>Bos taurus</i>	Cow	AID	NM_001038682	Verma and Aitken, 2005
<i>Ovis aries</i>	Sheep	A3Z1	EU864541	This study
<i>Ovis aries</i>	Sheep	A3Z2-Z3	DQ974646	Jónsson <i>et al.</i> , 2006
<i>Ovis aries</i>	Sheep	A3Z2	EU864542	This study
<i>Ovis aries</i>	Sheep	A3Z3	EU864543	This study
<i>Ovis aries</i>	Sheep	AID	EE793762	This study
<i>Sus scrofa</i>	Pig	A3Z2-Z3	DQ974646	Jónsson <i>et al.</i> , 2006
<i>Sus scrofa</i>	Pig	A3Z2	EU864539	This study
<i>Sus scrofa</i>	Pig	A3Z3	EU864540	This study
<i>Sus scrofa</i>	Pig	AID	BP157753	This study
<i>Tayssu tajacu</i>	Peccary	A3Z2-Z3	EU864537	This study
<i>Tayssu tajacu</i>	Peccary	AID	EU864538	This study
<i>Equus caballus</i>	Horse	A3Z1	XM_001499871	Genome project ID: 19129
<i>Equus caballus</i>	Horse	A3Z2	XM_001501833	Genome project ID: 19129
<i>Equus caballus</i>	Horse	A3Z3	XM_001501833	Genome project ID: 19129
<i>Felis catus</i>	Cat	A3Z2b-Z3	EF173021	Munk <i>et al.</i> , 2007
<i>Canis lupus</i>	Dog	A3Z1	XM_847690	Genome project ID: 12384
<i>Canis lupus</i>	Dog	A3Z2	AACN010393938	Kirkness <i>et al.</i> , 2003
<i>Canis lupus</i>	Dog	A3Z3	XM_538369	Genome project ID: 12384
<i>Canis lupus</i>	Dog	AID	NM_001003380	Ohmori <i>et al.</i> , 2004
<i>Mus musculus</i>	Mouse	A3Z2-Z3	NM_030255	Many; LaRue <i>et al.</i> , 2008
<i>Mus musculus</i>	Mouse	AID	NM_009645	Many; LaRue <i>et al.</i> , 2008
<i>Rattus norvegicus</i>	Rat	A3Z2-Z3	NM_001033703	Strausberg <i>et al.</i> , 2002
<i>Rattus norvegicus</i>	Rat	AID	XM_001060382	Twigger <i>et al.</i> , 2007
<i>Homo sapiens</i>	Human	A3A (A3Z1a)	NM_145699	Many; LaRue <i>et al.</i> , 2008
<i>Homo sapiens</i>	Human	A3B (A3Z2a-Z1b)	NM_004900	Many; LaRue <i>et al.</i> , 2008
<i>Homo sapiens</i>	Human	A3C (A3Z2b)	NM_014508	Many; LaRue <i>et al.</i> , 2008
<i>Homo sapiens</i>	Human	A3DE (A3Z2c-Z2d)	NM_152426	Many; LaRue <i>et al.</i> , 2008
<i>Homo sapiens</i>	Human	A3F (A3Z2e-Z2f)	NM_145298	Many; LaRue <i>et al.</i> , 2008
<i>Homo sapiens</i>	Human	A3G (A3Z2g-Z1c)	NM_021822	Many; LaRue <i>et al.</i> , 2008
<i>Homo sapiens</i>	Human	A3H (A3Z3)	NM_181773	Many; LaRue <i>et al.</i> , 2008
<i>Homo sapiens</i>	Human	AID	NM_020661	Many; LaRue <i>et al.</i> , 2008
<i>Pan troglodytes</i>	Chimp	A3H	DQ408607	OhAinle <i>et al.</i> , 2006
<i>Pan troglodytes</i>	Chimp	AID	NM_001071809	Zhou <i>et al.</i> , 2005
<i>Macaca mulatta</i>	Macaque	A3H	XM_001096739	Genome project ID: 16397
<i>Macaca mulatta</i>	Macaque	AID	XM_001113641	Genome project ID: 16397
<i>Hylobates syndactylus</i>	Gibbon	A3H	DQ408608	OhAinle <i>et al.</i> , 2006
<i>Pongo pygmaeus</i>	Orangutan	A3H	DQ408610	OhAinle <i>et al.</i> , 2006
<i>Gorilla gorilla</i>	Gorilla	A3H	DQ408609	OhAinle <i>et al.</i> , 2006
<i>Pan paniscus</i>	Bonobo	A3H	DQ408606	OhAinle <i>et al.</i> , 2006
<i>Miopithecus talapoin</i>	Talapoin	A3H	DQ408613	OhAinle <i>et al.</i> , 2006
<i>Papio anubis</i>	Baboon	A3H	DQ408605	OhAinle <i>et al.</i> , 2006
<i>Cercocebus atys</i>	Sooty mangabey	A3H	DQ408611	OhAinle <i>et al.</i> , 2006
<i>Saguinus labiatus</i>	Tamarin	A3H	DQ408614	OhAinle <i>et al.</i> , 2006
<i>Ateles belzebuth</i>	Spider	A3H	DQ408612	OhAinle <i>et al.</i> , 2006
<i>Lagothrix lagotricha</i>	Woolly monkey	A3H	DQ408615	OhAinle <i>et al.</i> , 2006

**Table S2.** Primers used to identify expressed *APOBEC3* transcripts from cow, sheep and pig PBMCs.

	<b>Outer 5' RACE</b>	<b>Inner 5' RACE</b>	<b>Outer 3' RACE</b>	<b>Inner 3' RACE</b>
Cow A3Z1	ACTCTGATGGCACCCTAAAAC	GCCTCTCCATCTCGTAGCAC	TTTGATCAACCGGAGAAAC	TCCTGAAGGAGAACCACCAC
Cow A3Z2-Z3	ATGCTCAGCGTCACATTCTG	GGAGGCACGTGAGTGGTATT	ACCAGCTGAAGCAGCGTAAT	GCATAAGACGAAGGCTCCAG
Cow A3Z2	Same as cow A3Z2-Z3	Same as cow A3Z2-Z3	CAGAATGTGACGCTGAGCAT	GGAGCCCATGTGGACATTAT
Cow A3Z3	n.a.	n.a.	Same as cow A3Z2-Z3	Same as cow A3Z2-Z3
Sheep A3Z1	n.a.	n.a.	TCCGTTCTTGAATCTGGAC	GAAGGAGAACCGCCACATAA
Sheep A3Z2-Z3	TCAGCGTCACATTCTGGTACA	AGTCCCAGCATAGACCTGGTT	AACCAGGTCTATGCTGGGACT	CTGGGGATGTACCAGAATGTG
Sheep A3Z2	Same as sheep A3Z2-Z3	Same as sheep A3Z2-Z3	Same as sheep A3Z2-Z3	Same as sheep A3Z2-Z3
Sheep A3Z3	n.a.	n.a.	Same as sheep A3Z2-Z3	Same as sheep A3Z2-Z3
Pig A3Z2-Z3	n.a.	n.a.	CCAAGGAGCTGGTTGATTC	CTGGAGCAATACAGCGAGAG
Pig A3Z2	n.a.	n.a.	ACGTCACCTGGTTCATCTCC	ACTGCTGGAACAACCTTCGTG
Pig A3Z3	n.a.	n.a.	Same as pig A3Z2-Z3	Same as pig A3Z2-Z3

n.a. = not applicable

## Paper IV



# The Restriction of Zoonotic PERV Transmission by Human APOBEC3G

Stefán R. Jónsson<sup>1,2,3,4\*</sup>, Rebecca S. LaRue<sup>1,2,3\*</sup>, Mark D. Stenglein<sup>1,2,3</sup>, Scott C. Fahrenkrug<sup>3,5</sup>, Valgerdur Andrésdóttir<sup>4</sup>, Reuben S. Harris<sup>1,2,3\*</sup>

**1** Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, Minnesota, United States of America, **2** Institute for Molecular Virology, University of Minnesota, Minneapolis, Minnesota, United States of America, **3** Arnold and Mabel Beckman Center for Transposon Research, University of Minnesota, Minneapolis, Minnesota, United States of America, **4** Institute for Experimental Pathology, University of Iceland, Reykjavik, Iceland, **5** Department of Animal Sciences, University of Minnesota, St. Paul, Minnesota, United States of America

The human APOBEC3G protein is an innate anti-viral factor that can dominantly inhibit the replication of some endogenous and exogenous retroviruses. The prospects of purposefully harnessing such an anti-viral defense are under investigation. Here, long-term co-culture experiments were used to show that porcine endogenous retrovirus (PERV) transmission from pig to human cells is reduced to nearly undetectable levels by expressing human APOBEC3G in virus-producing pig kidney cells. Inhibition occurred by a deamination-independent mechanism, likely after particle production but before the virus could immortalize by integration into human genomic DNA. PERV inhibition did not require the DNA cytosine deaminase activity of APOBEC3G and, correspondingly, APOBEC3G-attributable hypermutations were not detected. In contrast, over-expression of the sole endogenous APOBEC3 protein of pigs failed to interfere significantly with PERV transmission. Together, these data constitute the first proof-of-principle demonstration that APOBEC3 proteins can be used to fortify the innate anti-viral defenses of cells to prevent the zoonotic transmission of an endogenous retrovirus. These studies suggest that human APOBEC3G-transgenic pigs will provide safer, PERV-less xenotransplantation resources and that analogous cross-species APOBEC3-dependent restriction strategies may be useful for thwarting other endogenous as well as exogenous retrovirus infections.

Citation: Jónsson SR, LaRue RS, Stenglein MD, Fahrenkrug SC, Andrésdóttir V, et al (2007) The Restriction of Zoonotic PERV Transmission by Human APOBEC3G. PLoS ONE 2(9): e893. doi:10.1371/journal.pone.0000893

## INTRODUCTION

Viral zoonoses have impacted human populations through the ages. The 1918 influenza epidemic, the 2003 SARS incident and the ongoing HIV/AIDS pandemic constitute clear examples of virus transfer from birds [1], bats [2] and chimpanzees [3]. Domesticated animals can also function as zoonotic intermediates (e.g., [4,5]). Additional and unprecedented opportunities for zoonoses occur when live cells, tissues or organs are transplanted from one species to another [6]. However, despite risks and technical and immunological challenges, several xenotransplantation procedures have shown preclinical promise for treating diabetes, heart, kidney and other human diseases (e.g., [7–9]).

Pigs are favourable xenotransplantation sources because of their human-like physiology, large litters, short gestation period and genetic malleability [10]. However, pig to human virus transmission has been a concern since it was shown that porcine endogenous retroviruses (PERVs) could infect human cells in culture [6,11–13]. Although PERV transmission has yet to be documented in xenotransplantation patients, significant concerns still exist regarding PERV and other potentially pathogenic viruses [14,15]. Strategies to reduce the likelihood of PERV transmission have been proposed, such as selective breeding for lower levels of PERV, RNAi transgenesis to knock-down PERV expression or systematic deletion of active PERV copies (e.g., [15,16]). The first two are unlikely to be completely effective or risk-free and the third, albeit theoretically feasible, may be overly technical and prohibitively expensive. Therefore, alternative, robust and cost-effective methods to reduce PERV transmission and possible xenozoonotic infections are desirable.

APOBEC3G is a single-strand DNA cytosine deaminase best understood as a potent inhibitor of HIV-1 replication [17–23]. It can however also inhibit a variety of other exogenous and endogenous retroviruses/elements (e.g., [17,18,24–29]). APOBEC3G engages an assembling retrovirus particle, accesses the

RNA genome-containing virus core and, upon reverse transcription, deaminates cDNA cytosines to uracils (C-to-U). Catastrophic levels of uracil either directly inactivate the coding capacity of the virus or trigger the degradation of the viral DNA. The former manifests as genomic strand-specific guanine to adenine (G-to-A) hypermutations (cDNA strand C-to-T transitions). However, in several instances, it is noteworthy that the deaminase activity of APOBEC3G or other APOBEC3 proteins is partly or even completely dispensable (e.g., HIV-1, hepatitis B virus, L1 and Alu [27–32]).

Interestingly, throughout evolution, the retroviruses of many mammals appear to have become largely immune to APOBEC3G or to the APOBEC3G-like proteins of their hosts. HIV-1 expresses

**Academic Editor:** Douglas Nixon, University of California at San Francisco, United States of America

**Received:** July 30, 2007; **Accepted:** August 13, 2007; **Published:** September 12, 2007

**Copyright:** © 2007 Jónsson et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** These studies were funded in part by grants from the NIH to RSH and from Spring Point Project to SCF. SRJ was supported in part by the Icelandic Research Fund for Graduate Students and a 2004–2005 Val Bjornson Icelandic Exchange Scholarship. MDS was supported in part by a 3M Science and Technology Graduate Fellowship. RSH was supported in part by a Searle Scholarship and a University of Minnesota McKnight Land Grant Professorship. Sponsors did not contribute to the study design, to the collection, analysis and interpretation of data, to writing the paper or to the decision to submit it for publication.

**Competing Interests:** The authors have declared that no competing interests exist.

\* To whom correspondence should be addressed. E-mail: rsh@umn.edu

† These authors contributed equally to this work.

an accessory protein, Vif, which neutralizes APOBEC3G through ubiquitination and degradation [33–36]. Simian immunodeficiency virus (SIV) uses a similar Vif-dependent mechanism [17,26]. Foamy viruses employ an unrelated viral protein called Bet, for which the precise neutralization mechanism is currently unclear [37,38]. Murine leukaemia virus (MLV) and human T-lymphotropic virus-1 (HTLV-1) may simply avoid APOBEC3 proteins by preventing encapsidation [26,39,40].

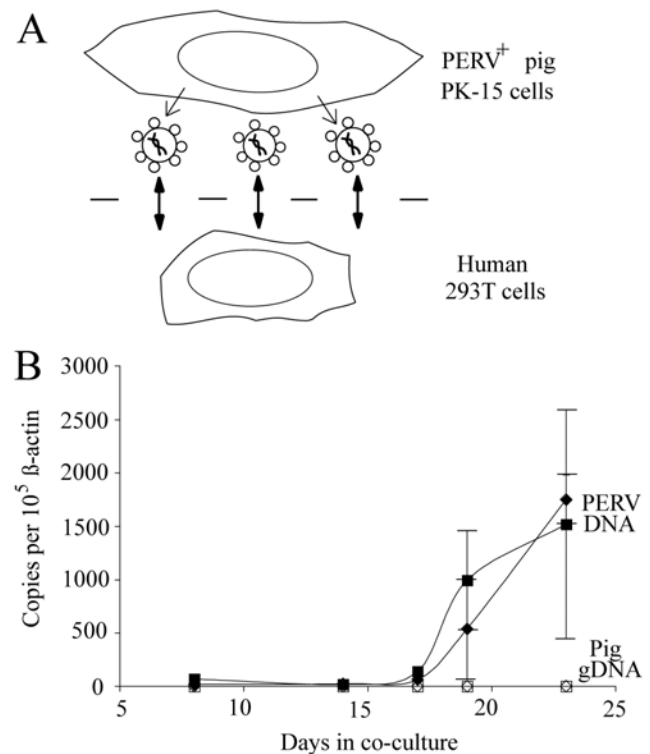
However, cell-based studies have indicated that an APOBEC3 protein from a mammal to which the virus has not yet adapted may provide an effective strategy for thwarting species-specific viral counter-defenses. For instance, human APOBEC3G can potentially inhibit the replication of SIV (except isolates such as SIV<sub>cpz</sub> which encode a Vif protein closely related to that of HIV-1), feline foamy virus and MLV (*e.g.*, [17–20,25,26,39,41,42]). Similarly, mouse APOBEC3 can potentially block HIV-1 replication (regardless of Vif), although it is completely unable to impede the replication of MLV [26,41]. One of the most dramatic examples to date used human and mouse APOBEC3 proteins to inhibit the mobilization of the yeast retrotransposons Ty1 and Ty2 [43,44]. Thus, it is reasonable to hypothesize that cross-species expression of an APOBEC3 protein may be used to create a powerful barrier to impede or perhaps even block retrovirus infection. Here, this rationale is applied to the specific question of whether human APOBEC3G expression can inhibit the transmission of PERV from pig to human cells. The results demonstrate that PERV transmission can be strongly inhibited by APOBEC3G.

## RESULTS

### A Co-culture Assay to Monitor PERV Transmission

To determine whether expression of human APOBEC3G would inhibit the transfer of PERV from pig to human cells it was first necessary to establish a long-term co-culture system. A trans-well assay was set up to monitor PERV transmission from pig kidney PK-15 fibroblasts to recipient human embryonic kidney 293T cells (Figure 1A). These two cell types were used because transmission from PK-15 to 293 cells had been reported previously [11]. The trans-well system enabled co-cultures to be sustained for several weeks, and it facilitated the recovery of each cell type for downstream analyses. An additional benefit of this co-culture system (not provided by transient assays) is that it enables the simultaneous analysis of multiple, endogenous PERV elements, which are precisely the targets one would want to monitor and ideally inhibit in (xeno)transplantation procedures.

At each co-culture passage point, surplus human 293T cells were used to prepare genomic DNA. PERV transmission was monitored by subjecting these samples to quantitative (Q)-PCR. PERV-specific *pol* gene PCR products could be detected in the human genomic DNA samples after approximately two weeks of continuous co-culture (Figure 1B). From the point of first detection onward, the total number of PERV transmissions continued to increase, averaging 190 new events per day per 50,000 cells ( $10^5$  *beta-actin* copies; SEM = 62; *n* = 5 experiments). Importantly, pig cells did not breach the 293T cell compartment because Q-PCR analyses of the same human genomic DNA samples failed to detect a concomitant transfer of pig genomic DNA (Figure 1B; also see Online Figure S1 for PERV *pol* gene Q-PCR standard curves, representative PERV-specific datasets and human *beta-actin* controls). Moreover, PERV copy number did not increase over a two week interval when infected 293T cells were grown in isolation, indicating that PERV was not replicating in the human cells and that the majority of the observed transmission events were derived from the PK-15 cells (*i.e.*, new events). These results



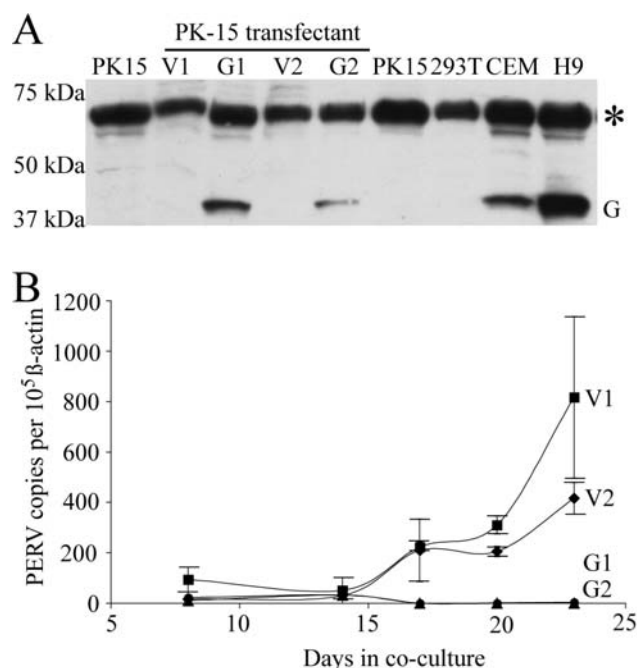
**Figure 1. PERV Transmission Assay.** (A) Schematic of the co-culture system. PERV transmitting PK-15 cells are grown on top of the membrane of the insert and human 293T cells on the bottom of the well of the culture dish. Virus particles are depicted diffusing through membrane pores. (B) The zoonotic transmission of PERV from PK-15 cells to 293T cells is shown by the time-dependent accumulation of PERV *pol* gene DNA in the human cells (solid diamonds and squares). No concomitant transfer of pig genomic DNA occurred through the duration of these experiments (open diamonds and squares). This graph summarizes data for two independently derived PK-15 clones, V1 (squares) and V2 (diamonds). All data points were calculated using results from duplicate Q-PCR reactions of genomic DNA from three parallel (but independent) co-cultures. The error bars indicate the SEM. See the Materials and Methods and Online Figure S1 for additional details, representative raw data and controls. doi:10.1371/journal.pone.0000893.g001

combine to indicate that the trans-well assay provided a robust system for monitoring *bona fide* zoonotic PERV transmissions.

### Human APOBEC3G Inhibits PERV Transmission

The second key step in addressing our experimental question was isolating PK-15 clones that stably expressed human APOBEC3G. Clones expressing human APOBEC3G cDNA or an empty vector control were established in parallel (Figure 2A). Immunoblotting identified clones with APOBEC3G levels similar to those in known APOBEC3G-expressing human T cell lines CEM and H9, which are non-permissive for growth of Vif-deficient HIV-1 [19]. Although it was impossible to achieve a physiologic expression level, the comparative immunoblot at least ensured that the levels of APOBEC3G were equal or lower than those present in well-studied, non-permissive human T cell lines.

Co-culture experiments were set up to compare PERV transmission from two independently derived PK-15 clones expressing human APOBEC3G and two vector expressing controls. Remarkably, the human APOBEC3G expressing PK-15 clones showed levels of PERV transmission that were lower than the Q-PCR detection threshold of approximately 10 copies (Figure 2B;

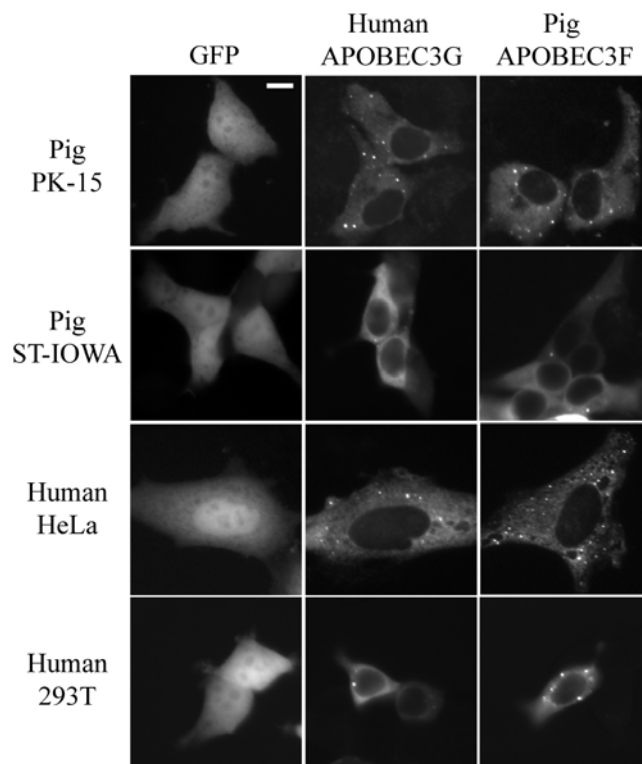


**Figure 2. Human APOBEC3G Inhibits PERV Transmission.** (A) An immunoblot showing PK-15 clones expressing human APOBEC3G (G1 and G2) or a control vector (V1 and V2). PK-15 and 293T cell lysates were used as negative controls. CEM and H9 were used as positive controls for APOBEC3G expression. A non-specific (but pan-species) band is shown as a protein loading control (marked by an asterisk). (B) Q-PCR data using genomic DNA prepared from 293T cells co-cultured with two independently derived APOBEC3G expressing PK-15 clones (G1 and G2, circles and triangles, respectively) or two vector control clones (V1 and V2, diamonds and squares, respectively). The experimental parameters are identical to those used in Figure 1B. doi:10.1371/journal.pone.0000893.g002

Online Figure S1). In contrast, the control clones showed high levels of PERV transfer by co-culture day 17 and transmission events continued to accumulate through the duration of the experiment. The kinetics of PERV transmission were similar to those reported in Figure 1B (these results also contributed to the aforementioned transfer rate calculations). These data were further corroborated by additional experiments where PERV transmission was monitored simultaneously by Q-PCR and by reverse-transcriptase ELISA assays (Online Figure S2).

### The Sub-cellular Distribution of Human APOBEC3G Is Virtually Identical in Human and Pig Cell Lines

An ultimate application of the technology described here raises the potential problem that human APOBEC3G may not be subjected to (proper) post-translational regulation in pig cells and it may therefore promote carcinogenesis. Expression of human APOBEC3G in a heterologous system has been shown to trigger elevated levels of genomic C/G-to-T/A transition mutation [44]. Therefore, to help mitigate this risk (in addition to establishing clones that expressed relatively modest APOBEC3G levels; above), we asked whether the predominantly cytoplasmic localization pattern of human APOBEC3G would be maintained in PK-15 and in a swine testes cell line, ST-IOWA (Figure 3; compare with other APOBEC3G reports [18,35,41]). Unlike some other APOBEC3 proteins such as human APOBEC3B, which is mostly nuclear, both human APOBEC3G and pig APOBEC3F appeared predominantly cytoplasmic in either the human or the pig cell lines (Figure 3; [28,31,41]). Both proteins



**Figure 3. The Sub-cellular Distribution of Human APOBEC3G in Human and Pig Cell Lines.** Sub-cellular distributions of GFP, human APOBEC3G-GFP and pig APOBEC3F-GFP in the indicated live pig and human cell lines. The scale bar in the top left panel indicates 10  $\mu$ m. doi:10.1371/journal.pone.0000893.g003

also appeared to concentrate in cytoplasmic punctae, which varied in number and were apparent in some of the cells regardless of species of origin (described previously for APOBEC3G; e.g., [45,46]). Overall, these near-identical localization patterns suggested that human APOBEC3G is not aberrantly regulated in pig cells and, interestingly, that these proteins might be subjected to the same cellular regulatory mechanism(s).

### PERV Appears Resistant to Porcine APOBEC3F

During the course of these experiments, we reported some of the activities of pig APOBEC3F [41]. It could strongly inhibit the replication of HIV (regardless of Vif) and modestly inhibit the replication of MLV, a gamma-retrovirus phylogenetically related to PERV. Therefore, we wondered whether pig APOBEC3F was expressed in PK-15 and, if so, whether PERV resists this cellular defense. To begin to address this possibility, RT-PCR was used to test PK-15 cells for pig APOBEC3F expression. Pig *APOBEC3F* mRNA was detected readily (Online Figure S3A). Full cDNA sequencing revealed that the predicted APOBEC3F protein of PK-15 cells was 98% identical to the variant we reported previously [41]. Eight amino acid differences were found, but both the PK15 and the previously reported APOBEC3F sequences were represented in pig genomic DNA sequences suggesting that these may be breed-specific polymorphisms (R.S.L. and R.S.H., manuscript in preparation). These observations indicated that either PERV resists the endogenous APOBEC3F protein of its host or that the level of suppression by pig APOBEC3F is not sufficient to inhibit PERV transmission.

To begin to distinguish between these two hypotheses, PK-15 clones over-expressing pig APOBEC3F were established and used

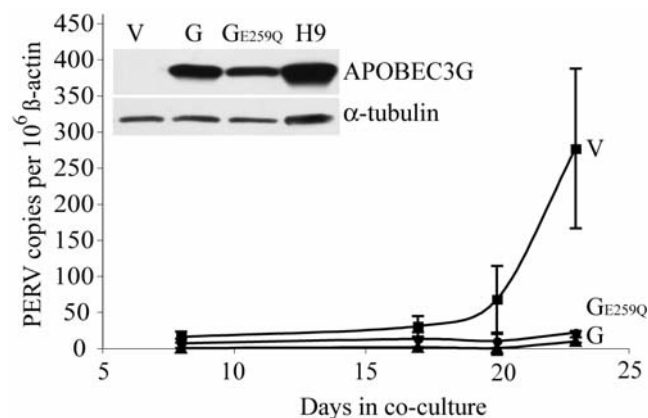
in transmission experiments. The former hypothesis was favored because pig APOBEC3F over-expression did not significantly interfere with PERV transmission (Online Figure S3B). These results were further supported by PCR experiments showing that PERV could be amplified readily from 293T cells that had been co-cultured with PK-15 over-expressing pig APOBEC3F (unlike the APOBEC3G scenario; below).

We further noted that it is highly unlikely that another resident APOBEC3 protein contributes to PERV restriction, because genomic DNA sequencing showed that pigs have only one *APOBEC3* gene, *APOBEC3F* (R.S.L. and R.S.H., manuscript in preparation). These observations combined to indicate that PERV is resistant to the endogenous APOBEC3 protein of its host. In hindsight, this was not particularly surprising given the emerging trend that (successful) retroviruses are selected in part by their ability to evade the APOBEC3 proteins of their host species (see Introduction).

### Evidence that Human APOBEC3G Inhibits PERV By a Deamination-Independent Mechanism

The hallmark of APOBEC3G-dependent retrovirus restriction is plus-strand G-to-A hypermutation, which is caused by the deamination of minus-strand cDNA C-to-U during reverse transcription [17,18,20,24,25]. The deamination of cytosines within single-strand DNA requires glutamate 259 (E259) of APOBEC3G [47–49]. Based on homology to structurally defined deaminases, E259 likely functions by helping position the water molecule that ultimately initiates the deamination reaction by attacking the cytosine ring (as a hydroxide; reviewed by [17,23,50,51]).

To determine whether DNA deamination is required the APOBEC3G-dependent inhibition of PERV transmission, we established a new set of PK-15 clones expressing APOBEC3G, APOBEC3G E259Q or a vector control (Figure 4, inset). Surprisingly, both APOBEC3G and the E259Q derivative diminished PERV transmission to near background levels



**Figure 4. The APOBEC3G-dependent Inhibition of PERV Transmission Is Deamination-Independent.** PERV-specific Q-PCR data using genomic DNA prepared from 293T cells co-cultured with PK-15 clones expressing APOBEC3G (G; triangles), APOBEC3G-E259Q (GE259Q; circles) or empty vector (V; squares). Two datasets, each with an independent PK-15 clone in three replica co-culture wells, were collected in parallel and averaged for each data point. One standard error of the mean is shown. The experimental parameters are identical to those used in Figure 1B. The inset immunoblots show the APOBEC3G and  $\alpha$ -tubulin levels of representative PK-15 clones expressing the indicated construct. The human T cell line H9 provided a positive control for APOBEC3G expression.

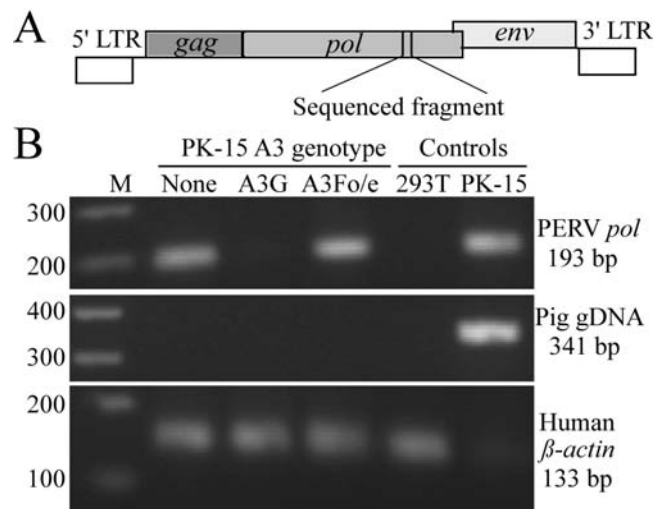
doi:10.1371/journal.pone.0000893.g004

(Figure 4). These data demonstrated that the mechanism of inhibition does not require the DNA deaminase activity of APOBEC3G. These data were further supported by the fact that plus strand G-to-A hypermutations were not apparent in the DNA of the rare transmission events that occurred in the presence of human APOBEC3G (below).

### Genetic Variation in Zoonosed PERV DNA Sequences

To begin to genotype the infectious PERVs and to further probe the mechanism of PERV restriction by human APOBEC3G, the PERV *pol* gene DNA was amplified from human 293T cells, cloned and sequenced (Figure 5A; Online Figure S4). Twenty-nine and twenty-two sequences were analysed from APOBEC3G and control experiments, respectively. To minimize possible PCR biases, any sequence that was recovered multiple times was considered one event, unless it arose from independent experiments. These DNA sequence analyses revealed several important points.

First, in contrast to vector control and pig APOBEC3F over-expressing co-cultures, PERV *pol* gene DNA was difficult to amplify from the genomic DNA of 293T cells that had been co-cultured with APOBEC3G-expressing PK-15 cells (Figure 5B and every significant sampling point in our Q-PCR experiments). Taking this together with the observation that APOBEC3G does not effect PK-15 virus production (similar RT levels were observed in cell-free supernatants in the presence or absence of APOBEC3G; data not shown), we infer that APOBEC3G restricts PERV transmission after virus production but before provirus integration (*i.e.*, between entry and integration). APOBEC3G may restrict PERV at an early reverse transcription stage, possibly by interfering with primer binding, DNA synthesis and/or integration



**Figure 5. Genetic Variation in Zoonosed PERV DNA Sequences.** (A) A schematic of the PERV genome showing coding regions (*gag*, *pol* and *env*) and long-terminal repeats (LTRs). The relevant 193 bp *pol* gene fragment is indicated. (B) An ethidium bromide-stained agarose gel showing that PERV *pol* gene DNA was amplified readily from 293T cells cultured with PK-15 control clones (None) and pig APOBEC3F over-expressing clones (A3Fo/e) but not with PK-15 clones expressing human APOBEC3G (A3G; top panel). PERV *pol* gene DNA (top panel) and pig genomic DNA (*APOBEC3F* locus; middle panel) PCR products were detected in PK-15 genomic DNA, whereas human *beta-actin* was strongly detected in the 293T cell genomic DNA samples (a much weaker amplification of pig *beta-actin* occurred because the human primers had partial complementarity to pig sequences, 20/21 and 17/18 nucleotides).

doi:10.1371/journal.pone.0000893.g005

as shown recently for APOBEC3G and HIV-1 substrates (*e.g.*, [30,52–55]).

Second, control co-culture PERV transmission events were exceptionally diverse, as 11 unique *pol* sequences were detected and only 4 were found multiple times (Online Figure S4). These data suggested that PK-15 cells have at least 11 active PERVs capable of infecting human 293T cells, a number consistent with previous studies that reported the existence of approximately 17–50 PERV copies in total (with only a fraction being replication-competent; [11,12]). In contrast, the rare PERV sequences derived from the APOBEC3G co-culture experiments showed a much lower genetic complexity. Only three unique sequences were recovered, each differing by a single nucleotide (Online Figure S4). In parallel experiments with HIV-based viruses, this APOBEC3G expression construct caused approximately 30 G-to-A hypermutations per 1000 bases analyzed (*e.g.*, [41]). Thus, approximately 12 G-to-A transitions should have been recovered in these PERV DNA analyses (nearly 90 if multiply recovered sequences would have been considered). The absence of hypermutated PERV proviral DNA provided further support for a deaminase-independent mechanism of restriction, which may share features with other instances described previously (*e.g.*, [27–30]).

## DISCUSSION

We have established a quantitative assay to monitor the zoonotic transmission of PERV to human 293T cells. Expression of human APOBEC3G in the pig PK-15 cell line strongly inhibited PERV zoonoses, while the endogenous APOBEC3F protein of pigs appeared considerably less effective. These data are the first to show that human APOBEC3G can inhibit PERV and the first to demonstrate that APOBEC3 proteins can be used purposefully to reduce if not prevent zoonotic retroviral infections. These results were not anticipated because human APOBEC3G has a relatively weak effect against the PERV-related gamma-retrovirus MLV [26,39].

Our data indicate that the engineering of pigs to express human APOBEC3G (or an equally potent non-porcine APOBEC3) may result in animals whose cells and tissues are much less likely to disseminate functional PERV. The deamination-independence of the restriction mechanism suggests that a catalytically inert APOBEC3G protein, such as E259Q, may be equally potent and simultaneously reduce the risk of cancer-promoting mutagenesis. APOBEC3G or APOBEC3G-E295Q expressing pigs may therefore constitute safer source animals for pig-to-human xenotransplantation procedures. In contrast to knockdown, knock-out (by gene targeting or selective breeding) or most chemical-based anti-viral approaches to neutralize PERV [15], the APOBEC3 antiviral defense system has several advantages including a potentially broad neutralizing activity (effective against PERV and likely several other endogenous and exogenous viruses) and an applicability to situations where many copies of a virus are already present in a genome. Analogous transgenic applications can be envisaged, such as using cross-species APOBEC3 expression to purposefully impede known viruses (*e.g.*, the AIDS virus HIV-1 or the Hepatitis B virus HBV). Moreover, for humans and other mammals with multiple APOBEC3 proteins, our data encourage the development of methods to induce/up-regulate endogenous APOBEC3 proteins, which have the capacity but may not normally restrict a particular virus (*e.g.*, human APOBEC3B and HIV-1).

## MATERIALS AND METHODS

### Cell lines, plasmids and co-culture experiments

The porcine kidney PK-15 fibroblast cell line and the swine testes ST-IOWA cell line were obtained from the ATCC and cultured in

Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% fetal bovine serum (Gemini), and 25 units/ml penicillin and 25 µg/ml streptomycin at 37°C and 5% CO<sub>2</sub>. Human embryonic kidney 293T and HeLa (A. Bielinsky, University of Minnesota) cell lines were grown under the same conditions. The T cell lines H9 and CEM (M. Malim, Kings College London) were cultured in RPMI-1640 supplemented with 10% fetal bovine serum (Gemini), and 25 units/ml penicillin and 25 µg/ml streptomycin at 37 °C and 5% CO<sub>2</sub>. Plasmids encoding human APOBEC3G, human APOBEC3G-E259Q and porcine APOBEC3F were described previously [41]. The human APOBEC3G and porcine APOBEC3F cDNA sequences used here are identical to GenBank accession numbers, NM\_021822 and NM\_001097446, respectively.

Stable APOBEC3G- or vector control-expressing PK-15 cell lines were constructed by transfection using FuGENE6 according to the manufacturer's protocol (Roche) or by electroporation (BioRad, 250V, 950 µF). Clones were selected using growth medium containing 1 mg/ml G418 (Roche), and APOBEC3G expressing clones were identified by immunoblotting using a polyclonal antibody toward human APOBEC3G (J. Lingappa, University of Washington). All PK-15 clones were maintained in growth medium supplemented with 250 µg/ml G418 to ensure stable expression.

Long-term co-culture assays were performed in 6 well tissue culture plates with inserts (Transwell®, Corning Inc.). This system uses a membrane with 0.4 µm diameter pores, which keeps the two cell types separated physically but simultaneously allows diffusion of nutrients and small molecules including virus particles of approximately 0.1 µm (including PERV). Each experiment was initiated with 75,000 PK-15 cells (insert) and 75,000 293T cells (well) as illustrated (Figure 1A). At 72 hr intervals, each cell type was washed with PBS, subjected to mild trypsinization and diluted into 4 parts fresh growth medium. Excess 293T cells were used to prepare genomic DNA (Qiagen DNeasy kit).

The rate of PERV transfer was calculated using the *pol* gene levels from the last two data points (usually spanning a 3 day period). The difference between these levels represents PERV *pol* gene DNA that has accumulated per 100,000 human *beta-actin* gene copies (50,000 cells assuming that the 293T cell line has two *beta-actin* copies) per time period. Individual rates from 5 independent experiments were averaged to determine the overall transmission rate (190+/-62 events per day per 50,000 cells). Data from Figures 1B, 2B and S2A contributed to rate calculations.

### Genomic DNA Q-PCR assays

Genomic DNA was isolated from human 293T cells using the DNeasy kit (Qiagen). Duplicate 25 µl PCR reactions consisting of 10 ng of 293T genomic DNA, 100 nM primers and 2× iQ SYBR Green super mix (BioRad) were run on an iCycler iQ Multicolor Real-Time PCR detection System (BioRad). The thermocycler conditions consisted of an initial denaturation of 95°C for 5 min and 50 cycles of denaturation (95°C for 15 sec) and annealing (58°C for 30 sec). After the 50 cycles, a melting curve analysis (55°C to 95°C) was performed to confirm product specificity. The cycle threshold (C<sub>T</sub>) was generated using BioRad software and it was used to calculate the amount of target DNA (PERV *pol* or human *beta-actin*). A standard curve was generated using the method of Dorak [56] and a dilution series (10 to 10<sup>7</sup> copies) of a linearized plasmid containing the relevant 193 bp PERV *pol* gene fragment. The equation generated from the standard curve (slope and y intercept) was used to determine the efficiency of the PCR reaction and to quantify the number of PERV *pol* gene or human *beta-actin* copies in the Q-PCR reactions. PERV copy

numbers were normalized to those of *beta-actin* using the method of [56]. The primer sets used in this study were: PERV *pol* (193 bp): 5'-AAC CCT TTA CCC TTT ATG TGG AT and 5'-AAA GTC AAT TTG TCA GCG TCC TT; human *beta-actin* (133 bp)[57]: 5'-ATC ATG TTT GAG ACC TTC AA and 5'-AGA TGG GCA CAG TGT GGG T; pig genomic DNA (341 bp product specific to intron 5 of the pig *APOBEC3F* locus): 5'-TGG GGA GTG TGG AAT TAA CG and 5'-GGG GGT TAA GAA CCC AAC AT.

### RT-PCR experiments

Standard reverse transcription (RT)-PCR reactions were performed using RNA prepared from PK-15 cells (TRIzol protocol, Invitrogen), M-MLV reverse transcriptase was used for cDNA synthesis using an oligo dT primer (Ambion) and Taq polymerase was used for PCR (Roche). The primers specific to pig APOBEC3F were 5'-TGG TCA CAG AGC TGA AGC AG and 5'-TTG TTT TGG AAG CAG CCT TT (175 bp). The semi-nested primer set used to detect plasmid-expressed pig APOBEC3F was 5'-CCA AGG AGC TGG TTG ATT TC (exon 6, reaction 1), 5'-CTG GAG CAA TAC AGC GAG AG (exon 7, reaction 2) and 5'-TAG AAG GCA CAG TCG AGG, with the latter being vector specific (319 bp and 190 bp products, respectively). The mammalian *beta-actin* primers were 5'-CCT TCA ATT CCA TCA TGA AGT G and 5'-CCA CAT CTG CTG GAA GGT (236 bp). These primers amplify equally well a 236 bp *beta-actin* fragment from all mammals tested, including pigs and humans (e.g., Online Figure S3).

### Fluorescent microscopy

The human APOBEC3G-GFP, pig APOBEC3F-GFP and GFP expression constructs were described previously [28,41]. The pig and human cell lines were maintained as above. One day prior to transfection, 5,000–20,000 cells were seeded onto LabTek chambered coverglasses (Nunc). After 24 hrs incubation, these cells were transfected with 250 ng of the relevant plasmid construct. After 24 hrs of additional incubation, images of the live cells were acquired using a Zeiss Axiovert 200 microscope at 400× total magnification. Images were analyzed using Image J software (<http://rsb.info.nih.gov/ij>).

### Reverse transcriptase activity assays

Whole cell protein extracts were prepared from 293T cells by suspending 500,000 cells in PBS, sonicating twice for 5 seconds and clarifying the lysates by centrifugation. Soluble protein levels were quantified using a BioRad Bradford assay. 10 µg of cell lysate was tested for reverse transcriptase activity using a C-type-RT activity assay (Cavidi Tech) following the manufacturers' instructions. Cell-free PK-15 supernatants (PERV-containing) were assayed directly using the Cavidi Tech ELISA assay.

### PERV *pol* gene DNA sequence analyses

Human 293T cell genomic DNA was prepared from terminal co-cultures and 50 ng was used for high fidelity, PERV *pol* gene-specific PCR reactions (Phusion polymerase; Finnzymes). 193 bp products were cloned using the Zero Blunt TOPO PCR Cloning kit (Invitrogen) and sequenced (University of Minnesota Advanced Genetic Analysis Facility). Sequence comparisons were performed using Sequencher software (Gene Codes Corp.) and publicly available Clustal W alignment algorithms (<http://align.genome.jp/>).

## SUPPORTING INFORMATION

**Figure S1** Quantitative Real-time PCR Analyses. (A) Standard curves depicting Q-PCR data obtained using dilutions of a linearized PERV *pol* gene plasmid alone (squares) or diluted plasmid plus 10 ng of 293T cell genomic DNA (diamonds). Under both conditions, all template amounts (10 to 10<sup>7</sup> copies) amplified efficiently (the log-linear slope equations are shown). The correlation co-efficiency value ( $R^2$ ), which reports the technical accuracy of the assay, is also indicated. The standard curve data points were the average of 2 independent reactions with deviations smaller than the symbols (i.e.,  $C^T$  errors for each point ranged from 0 to 0.4). (B) Two representative control Q-PCR datasets showing the amplification of PERV *pol* gene DNA from pig PK-15 cell genomic DNA (circles). Two additional control Q-PCR datasets showing that the PERV-specific primers fail to amplify product from uninfected human 293T cell genomic DNA (squares). The reaction threshold, 10 times the mean standard deviation of the background fluorescence level (BioRad), is indicated. (C) Representative co-culture Q-PCR amplification curves of PERV *pol* gene DNA. Template genomic DNA isolated from human 293T cells co-cultured with vector expressing PK-15 cells (diamonds) or human APOBEC3G-expressing PK-15 cells (triangles) was used. (D) Representative Q-PCR amplification curves of the 293T cell *beta-actin* gene, which served as an internal standard for quantifying the real-time PCR data. Raw Q-PCR data will be made available on request.

Found at: doi:10.1371/journal.pone.0000893.s001 (9.93 MB TIF)

**Figure S2** APOBEC3G inhibits PERV transmission. (A) A graph showing the accumulation of PERV *pol* gene-specific PCR products in 293T cells co-cultured with a control cell line (V3) but not with an APOBEC3G-expressing cell line (G1). The data points were an average of two Q-PCR runs and the difference between each run was smaller than the plotted symbol. The experimental parameters were identical to those used in the experiments shown in Figures 1B and 2B. (B) Relative levels of reverse transcriptase(RT)-activity detected in soluble extracts of day 28 co-cultured 293T cells, which were used to generate the Q-PCR data shown in Figure S2A. Uninfected 293T cell lysates had a relatively high endogenous RT activity. Therefore, to help with the presentation of these data, this level was normalized to one and all of the other data were calculated relative to this value. The level of RT activity in PK-15 extracts was much higher than that of 293T cell extracts (+/-PERV) and it had reached saturation (out of range) when these data were collected.

Found at: doi:10.1371/journal.pone.0000893.s002 (4.76 MB TIF)

**Figure S3** Pig APOBEC3F Is Expressed in PK-15 Cells and its Over-expression Does Not Markedly Inhibit PERV Transmission. (A) An image of an ethidium bromide-stained agarose gel showing the results of an RT-PCR amplification experiment using PK-15 cellular RNA and appropriate controls. The top panel shows that PK-15 and representative PK-15 derived clones all expressed pig APOBEC3F, as indicated by the specific 175 bp pig APOBEC3F PCR product (confirmed by DNA sequencing). 293T cell mRNA and a diluted pig APOBEC3F expression plasmid were used as negative and positive controls, respectively. A larger, non-specific band was apparent only in the 293T cell RT-PCR reactions. The bottom panel shows that a conserved, 236 bp *beta-actin* gene fragment could be amplified from both PK-15 cells and human 293T cells (but not from diluted plasmid DNA). Note that this primer set differs from the human-specific set used in the Q-PCR experiments. The sizes of the marker (M) DNA bands are shown. (B) An image of an ethidium bromide-stained agarose gel showing

expression of plasmid-derived pig APOBEC3F in PK-15 cells after 26 days of continuous co-culture. Non-transfected (NT) cells and diluted APOBEC3F plasmid DNA (pDNA) provided negative and positive controls, respectively. The larger 319 bp (far right lane only) and smaller 190 bp bands are the specific PCR products of the first and second rounds of semi-nested PCR, respectively (confirmed by DNA sequencing). (C) A histogram summarizing the level of PERV transmission that was observed after 23 days of co-culturing human 293T cells with PK-15 cells expressing a vector control or over-expressing pig APOBEC3F. Two datasets, each with an independent PK-15 clone in three replica co-culture wells, were collected in parallel and averaged for each histogram bar. One standard error of the mean is shown. The experimental parameters are identical to those used in Figure 1B.

Found at: doi:10.1371/journal.pone.0000893.s003 (8.52 MB TIF)

**Figure S4** Genetic Variation in Zoonosed PERV *pol* Gene Sequences. (A) Sequences of the PERV *pol* gene fragments cloned from 293T cells co-cultured with control vector-expressing PK-15 cells. The number of times that each sequence was recovered is shown (N). Experiments 1 and 2 used genomic DNA prepared from the 293T cells used to generate the data shown in Online Figure S2 (day 28 samples) and Figure 2B (day 23), respectively. The most frequently detected 147 bp PERV *pol* gene sequence is shown in its entirety (which together with PCR primers makes up

the 193 bp product shown in Figure 5). Identical nucleotides in other sequences are represented by dashes and non-identical nucleotides by the indicated DNA bases. GenBank accession numbers are shown for *pol* gene fragments with 100% identity to previously reported sequences. (B) Sequences of the PERV *pol* gene fragments cloned from 293T cells co-cultured with control APOBEC3G-expressing PK-15 cells. Parameters are identical to those described above.

Found at: doi:10.1371/journal.pone.0000893.s004 (0.05 MB DOC)

## ACKNOWLEDGMENTS

We thank Y. Blanchard, W. Brown, P. Hackett, E. Hendrickson, M. Hertzberg, L. Mansky, W. Mothes, M. Murtaugh and H. Schuurman for valuable feedback, M. Titus for microscopy facilities, J. Lingappa for antisera, G. Haché for assistance with cell culture, N. Somia for *beta-actin* primer sequences and M. Malim, A. Bielinsky and the ATCC for cell lines.

## Author Contributions

Conceived and designed the experiments: RH SJ RL MS. Performed the experiments: RH SJ RL MS. Analyzed the data: RH SJ SF VA RL MS. Contributed reagents/materials/analysis tools: SF VA. Wrote the paper: RH. Other: Contributed to the discussion: VA SF. Helped edit the manuscript: VA SF.

## REFERENCES

- Kash JC, Tumpey TM, Prohl SC, Carter V, Perwitasari O, et al. (2006) Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. *Nature* 443: 578–581.
- Li W, Shi Z, Yu M, Ren W, Smith C, et al. (2005) Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310: 676–679.
- Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, et al. (1999) Origin of HIV-1 in the chimpanzee *Pan troglodytes* troglodytes. *Nature* 397: 436–441.
- AbuBakar S, Chang LY, Ali AR, Sharifah SH, Yusoff K, et al. (2004) Isolation and molecular identification of Nipah virus from pigs. *Emerg Infect Dis* 10: 2228–2230.
- Lipatov AS, Govorkova EA, Webby RJ, Ozaki H, Peiris M, et al. (2004) Influenza: emergence and control. *J Virol* 78: 8951–8959.
- Weiss RA (1999) Xenografts and retroviruses. *Science* 285: 1221–1222.
- Hering BJ, Wijkstrom M, Graham ML, Hardested M, Aasheim TC, et al. (2006) Prolonged diabetes reversal after intraportal xenotransplantation of wild-type porcine islets in immunosuppressed nonhuman primates. *Nat Med* 12: 301–303.
- Kuwaki K, Tseng YL, Dor FJ, Shimizu A, Houser SL, et al. (2005) Heart transplantation in baboons using alpha1,3-galactosyltransferase gene-knockout pigs as donors: initial experience. *Nat Med* 11: 29–31.
- Yamada K, Yazawa K, Shimizu A, Iwanaga T, Hisashi Y, et al. (2005) Marked prolongation of porcine renal xenograft survival in baboons through the use of alpha1,3-galactosyltransferase gene-knockout donors and the cotransplantation of vascularized thymic tissue. *Nat Med* 11: 32–34.
- Cooper DK, Gollackner B, Sachs DH (2002) Will the pig solve the transplantation backlog? *Annu Rev Med* 53: 133–147.
- Patience C, Takeuchi Y, Weiss RA (1997) Infection of human cells by an endogenous retrovirus of pigs. *Nat Med* 3: 282–286.
- Le Tissier P, Stoye JP, Takeuchi Y, Patience C, Weiss RA (1997) Two sets of human-tropic pig retrovirus. *Nature* 389: 681–682.
- Specke V, Rubant S, Denner J (2001) Productive infection of human primary cells and cell lines with porcine endogenous retroviruses. *Virology* 285: 177–180.
- Paradis K, Langford G, Long Z, Heneine W, Sandstrom P, et al. (1999) Search for cross-species transmission of porcine endogenous retrovirus in patients treated with living pig tissue. The XEN 111 Study Group. *Science* 285: 1236–1241.
- Takeuchi Y, Magre S, Patience C (2005) The potential hazards of xenotransplantation: an overview. *Rev Sci Tech* 24: 323–334.
- Dieckhoff B, Karlas A, Hofmann A, Kues WA, Petersen B, et al. (2007) Inhibition of porcine endogenous retroviruses (PERVs) in primary porcine cells by RNA interference using lentiviral vectors. *Arch Virol* 152: 629–634.
- Cullen BR (2006) Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors. *J Virol* 80: 1067–1076.
- Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, et al. (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* 424: 99–103.
- Sheehy AM, Gaddis NC, Choi JD, Malim MH (2002) Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 418: 646–650.
- Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, et al. (2003) The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* 424: 94–98.
- Holmes RK, Malim MH, Bishop KN (2007) APOBEC-mediated viral restriction: not simply editing? *Trends Biochem Sci* 32: 118–128.
- Bieniasz PD (2004) Intrinsic immunity: a front-line defense against viral attack. *Nat Immunol* 5: 1109–1115.
- Harris RS, Liddament MT (2004) Retroviral restriction by APOBEC proteins. *Nat Rev Immunol* 4: 868–877.
- Esnault C, Heidmann O, Delebecque F, Dewannieux M, Ribet D, et al. (2005) APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* 433: 430–433.
- Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, et al. (2003) DNA deamination mediates innate immunity to retroviral infection. *Cell* 113: 803–809.
- Mariani R, Chen D, Schrofelbauer B, Navarro F, Konig R, et al. (2003) Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* 114: 21–31.
- Sasada A, Takaori-Kondo A, Shirakawa K, Kobayashi M, Abudu A, et al. (2005) APOBEC3G targets human T-cell leukemia virus type 1. *Retrovirology* 2: 32.
- Stenglein MD, Harris RS (2006) APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J Biol Chem* 281: 16837–16841.
- Turelli P, Mangeat B, Jost S, Vianin S, Trono D (2004) Inhibition of hepatitis B virus replication by APOBEC3G. *Science* 303: 1829.
- Bishop KN, Holmes RK, Malim MH (2006) Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. *J Virol* 80: 8450–8458.
- Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, et al. (2006) Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci U S A* 103: 8780–8785.
- Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, et al. (2006) APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem* 281: 22161–22172.
- Yu X, Yu Y, Liu B, Luo K, Kong W, et al. (2003) Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* 302: 1056–1060.
- Coticello SG, Harris RS, Neuberger MS (2003) The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr Biol* 13: 2009–2013.
- Marin M, Rose KM, Kozak SL, Kabat D (2003) HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat Med* 9: 1398–1403.

36. Sheehy AM, Gaddis NC, Malim MH (2003) The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med* 9: 1404–1407.
37. Russell RA, Wiegand HL, Moore MD, Schafer A, McClure MO, et al. (2005) Foamy virus Bet proteins function as novel inhibitors of the APOBEC3 family of innate antiretroviral defense factors. *J Virol* 79: 8724–8731.
38. Lochelt M, Romen F, Bastone P, Muckenfuss H, Kirchner N, et al. (2005) The antiretroviral activity of APOBEC3 is inhibited by the foamy virus accessory Bet protein. *Proc Natl Acad Sci U S A* 102: 7982–7987.
39. Abudu A, Takaori-Kondo A, Izumi T, Shirakawa K, Kobayashi M, et al. (2006) Murine retrovirus escapes from murine APOBEC3 via two distinct novel mechanisms. *Curr Biol* 16: 1565–1570.
40. Derse D, Hill SA, Princler G, Lloyd P, Heidecker G (2007) Resistance of human T cell leukemia virus type 1 to APOBEC3G restriction is mediated by elements in nucleocapsid. *Proc Natl Acad Sci U S A* 104: 2915–2920.
41. Jonsson SR, Hache G, Stenglein MD, Fahrenkrug SC, Andresdottir V, et al. (2006) Evolutionarily conserved and non-conserved retrovirus restriction activities of artiodactyl APOBEC3F proteins. *Nucleic Acids Res* 34: 5683–5694.
42. Sawyer SL, Emerman M, Malik HS (2004) Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol* 2: E275.
43. Dutko JA, Schafer A, Kenny AE, Cullen BR, Curcio MJ (2005) Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases. *Curr Biol* 15: 661–666.
44. Schumacher AJ, Nissley DV, Harris RS (2005) APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. *Proc Natl Acad Sci U S A* 102: 9854–9859.
45. Wichroski MJ, Robb GB, Rana TM (2006) Human retroviral host restriction factors APOBEC3G and APOBEC3F localize to mRNA processing bodies. *PLoS Pathog* 2: e41.
46. Gallois-Montbrun S, Kramer B, Swanson CM, Byers H, Lynham S, et al. (2007) Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules. *J Virol* 81: 2165–2178.
47. Haché G, Liddament MT, Harris RS (2005) The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. *J Biol Chem* 280: 10920–10924.
48. Navarro F, Bollman B, Chen H, Konig R, Yu Q, et al. (2005) Complementary function of the two catalytic domains of APOBEC3G. *Virology* 333: 374–386.
49. Newman EN, Holmes RK, Craig HM, Klein KC, Lingappa JR, et al. (2005) Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr Biol* 15: 166–170.
50. Huthoff H, Malim MH (2005) Cytidine deamination and resistance to retroviral infection: towards a structural understanding of the APOBEC proteins. *Virology* 334: 147–153.
51. Chiu YL, Greene WC (2006) Multifaceted antiviral actions of APOBEC3 cytidine deaminases. *Trends Immunol* 27: 291–297.
52. Holmes RK, Koning FA, Bishop KN, Malim MH (2007) APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation. Comparisons with APOBEC3G. *J Biol Chem* 282: 2587–2595.
53. Guo F, Cen S, Niu M, Saadatmand J, Kleiman L (2006) Inhibition of formula-primed reverse transcription by human APOBEC3G during human immunodeficiency virus type 1 replication. *J Virol* 80: 11710–11722.
54. Luo K, Wang T, Liu B, Tian C, Xiao Z, et al. (2007) Cytidine deaminases APOBEC3G and APOBEC3F interact with human immunodeficiency virus type 1 integrase and inhibit proviral DNA formation. *J Virol* 81: 7238–7248.
55. Mbisa JL, Barr R, Thomas JA, Vandegraaff N, Dorweiler IJ, et al. (2007) Human Immunodeficiency Virus Type 1 cDNAs Produced in the Presence of APOBEC3G Exhibit Defects in Plus-Strand DNA Transfer and Integration. *J Virol* 81: 7099–7110.
56. Dorak MT (2006) Real-time PCR. New York: Taylor and Francis Group.
57. Agarwal S, Nikolai B, Yamaguchi T, Lech P, Somia NV (2006) Construction and use of retroviral vectors encoding the toxic gene barnase. *Mol Ther* 14: 555–563.

## Legends to the Supporting Online Figures

### Figure S1. Quantitative Real-time PCR Analyses

(A) Standard curves depicting Q-PCR data obtained using dilutions of a linearized PERV *pol* gene plasmid alone (squares) or diluted plasmid plus 10 ng of 293T cell genomic DNA (diamonds). Under both conditions, all template amounts ( $10$  to  $10^7$  copies) amplified efficiently (the log-linear slope equations are shown). The correlation coefficient value ( $R^2$ ), which reports the technical accuracy of the assay, is also indicated. The standard curve data points were the average of 2 independent reactions with deviations smaller than the symbols (*i.e.*,  $C_T$  errors for each point ranged from 0 to 0.4).

(B) Two representative control Q-PCR datasets showing the amplification of PERV *pol* gene DNA from pig PK-15 cell genomic DNA (circles). Two additional control Q-PCR datasets showing that the PERV-specific primers fail to amplify product from uninfected human 293T cell genomic DNA (squares). The reaction threshold, 10 times the mean standard deviation of the background fluorescence level (BioRad), is indicated.

(C) Representative co-culture Q-PCR amplification curves of PERV *pol* gene DNA. Template genomic DNA isolated from human 293T cells co-cultured with vector expressing PK-15 cells (diamonds) or human APOBEC3G-expressing PK-15 cells (triangles) was used.

(D) Representative Q-PCR amplification curves of the 293T cell  $\beta$ -*actin* gene, which served as an internal standard for quantifying the real-time PCR data. Raw Q-PCR data will be made available on request.

**Figure S2. APOBEC3G inhibits PERV transmission**

(A) A graph showing the accumulation of PERV *pol* gene-specific PCR products in 293T cells co-cultured with a control cell line (V3) but not with an APOBEC3G-expressing cell line (G1). The data points were an average of two Q-PCR runs and the difference between each run was smaller than the plotted symbol. The experimental parameters were identical to those used in the experiments shown in Figures 1B and 2B.

(B) Relative levels of reverse transcriptase(RT)-activity detected in soluble extracts of day 28 co-cultured 293T cells, which were used to generate the Q-PCR data shown in Figure S2A. Uninfected 293T cell lysates had a relatively high endogenous RT activity. Therefore, to help with the presentation of these data, this level was normalized to one and all of the other data were calculated relative to this value. The level of RT activity in PK-15 extracts was much higher than that of 293T cell extracts (+/- PERV) and it had reached saturation (out of range) when these data were collected.

**Figure S3. Pig APOBEC3F Is Expressed in PK-15 Cells and Its Over-expression Does Not Markedly Inhibit PERV Transmission**

(A) An image of an ethidium bromide-stained agarose gel showing the results of an RT-PCR amplification experiment using PK-15 cellular RNA and appropriate controls. The top panel shows that PK-15 and representative PK-15 derived clones all expressed pig APOBEC3F, as indicated by the specific 175 bp pig APOBEC3F PCR product (confirmed by DNA sequencing). 293T cell mRNA and a diluted pig APOBEC3F expression plasmid were used as negative and positive controls, respectively. A larger, non-specific band was apparent only in the 293T cell RT-PCR reactions. The bottom

panel shows that a conserved, 236 bp *β-actin* gene fragment could be amplified from both PK-15 cells and human 293T cells (but not from diluted plasmid DNA). Note that this primer set differs from the human-specific set used in the Q-PCR experiments. The sizes of the marker (M) DNA bands are shown.

(B) An image of an ethidium bromide-stained agarose gel showing expression of plasmid-derived pig APOBEC3F in PK-15 cells after 26 days of continuous co-culture. Non-transfected (NT) cells and diluted APOBEC3F plasmid DNA (pDNA) provided negative and positive controls, respectively. The larger 319 bp (far right lane only) and smaller 190 bp bands are the specific PCR products of the first and second rounds of semi-nested PCR, respectively (confirmed by DNA sequencing).

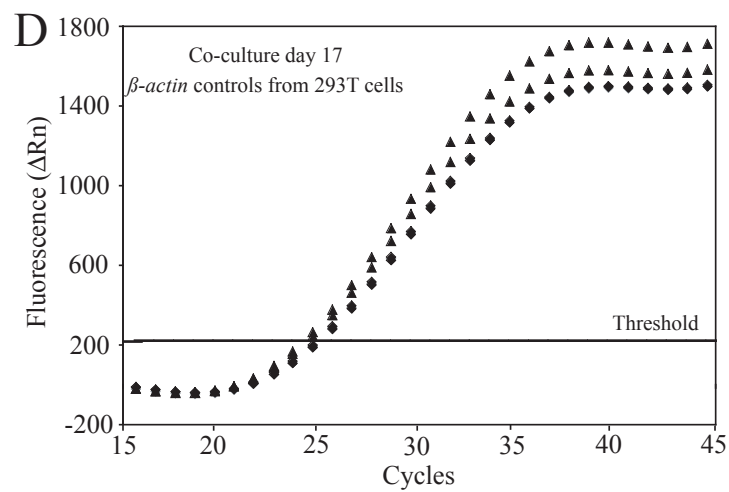
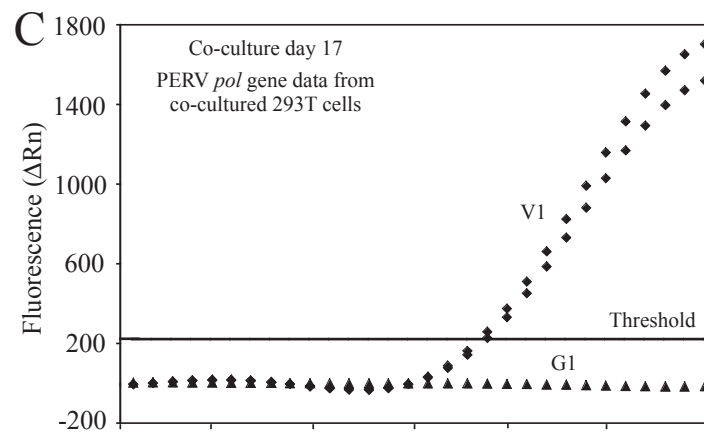
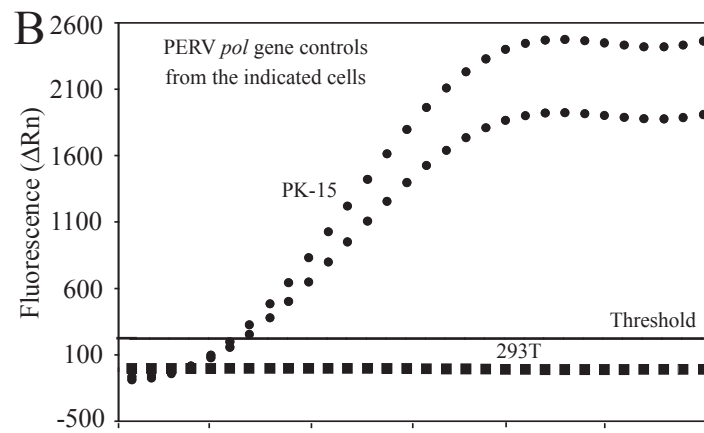
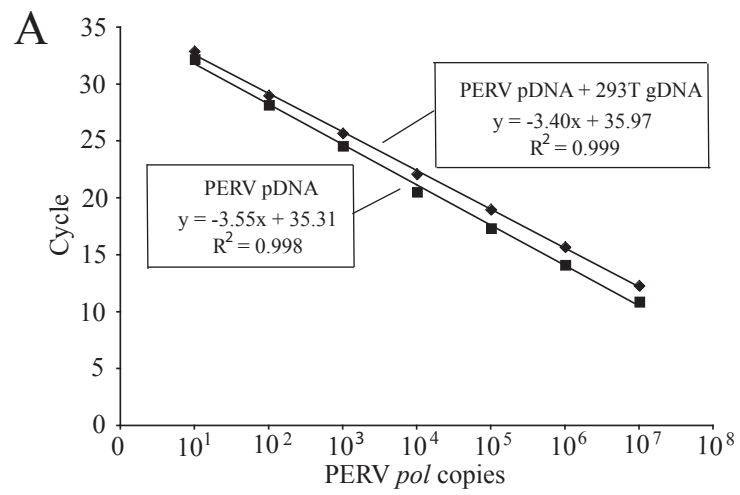
(C) A histogram summarizing the level of PERV transmission that was observed after 23 days of co-culturing human 293T cells with PK-15 cells expressing a vector control or over-expressing pig APOBEC3F. Two datasets, each with an independent PK-15 clone in three replica co-culture wells, were collected in parallel and averaged for each histogram bar. One standard error of the mean is shown. The experimental parameters are identical to those used in Figure 1B.

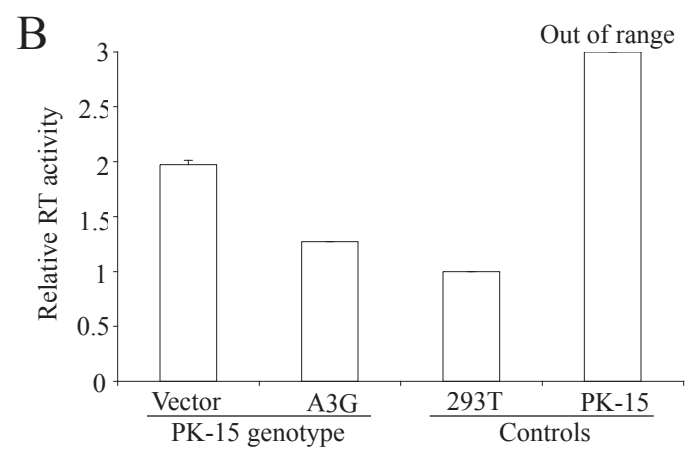
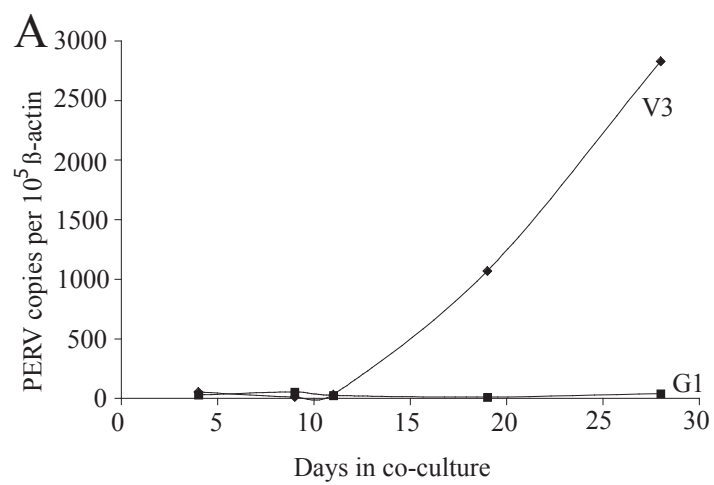
**Figure S4.** Genetic Variation in Zoonosed PERV *pol* Gene Sequences

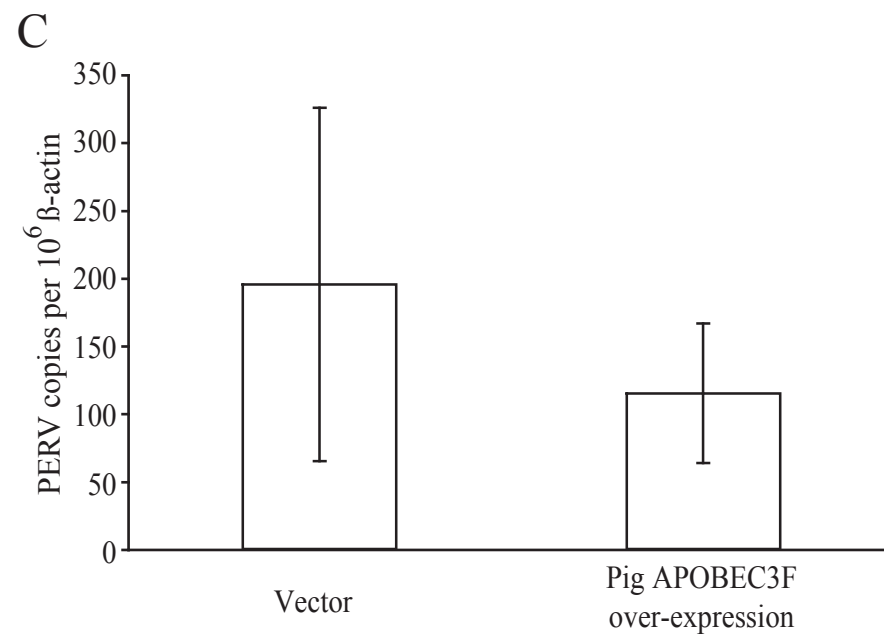
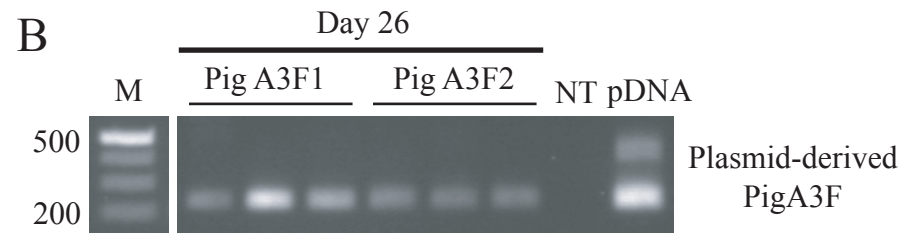
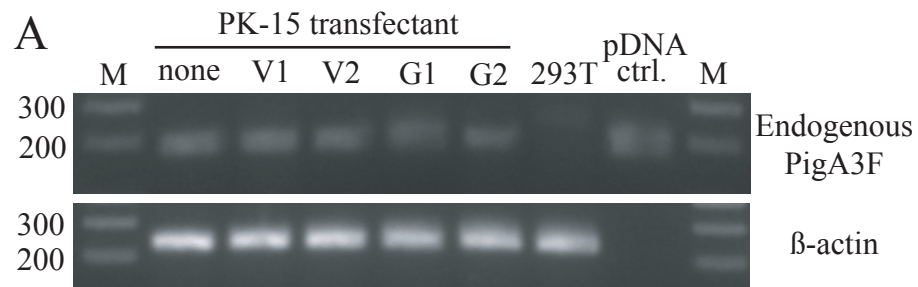
(A) Sequences of the PERV *pol* gene fragments cloned from 293T cells co-cultured with control vector-expressing PK-15 cells. The number of times that each sequence was recovered is shown (N). Experiments 1 and 2 used genomic DNA prepared from the 293T cells used to generate the data shown in Online Figure S2 (day 28 samples) and Figure 2B (day 23), respectively. The most frequently detected 147 bp PERV *pol* gene

sequence is shown in its entirety (without the PCR primers, which together make up the 193 bp product shown in Figure 5). Identical nucleotides in other sequences are represented by dashes and non-identical nucleotides by the indicated DNA bases. GenBank accession numbers are shown for *pol* gene fragments with 100% identity to previously reported sequences. PERV DNA sequences detected more than once are asterisked in Figure 5C.

(B) Sequences of the PERV *pol* gene fragments cloned from 293T cells co-cultured with control APOBEC3G-expressing PK-15 cells. Parameters are identical to those described above.







## A PERV transfers in vector control experiments

<u>N</u>	<u>Expt</u>	<u>Sequence</u>	<u>GB #</u>
5	1	GACCAGTATGGCCACAGCTGCGATAGCCTTCAGACATACGGGCCAACCACTGGCTACAGGATCAAGCTTCTTTGACAGGTAGGCAACAGGTCTCCTCCATGGTCCTAGGGTTTGGGTTAAAACTCCTCGGGCTACTCCCTTACGCTC	AY099323
3	2	-----G-----G-----C-----	
2	1	-----T-----G-----G-----C-----C-----T--	
1	1	-----G-----G-----A-----C-----	
1	1	-----G-----G-----A-----C-----	
1	1	-----G-----G-----A-----C-----	
1	1	-----G-----G-----A-----C-----	
1	1	-----G-----G-----A-----C-----	
1	1	-----G-----G-----A-----C-----	
1	1	-----G-----G-----A-----C-----	
1	1	-----G-----G-----A-----C-----	
1	1	-----G-----G-----A-----C-----	
1	1	-----G-----G-----A-----C-----	
1	1	-----G-----G-----A-----C-----	
1	1	-----G-----G-----A-----C-----	
1	1	-----G-----G-----A-----C-----	
1	2	-----C-----G-----T-----T-----G-----A-----T-----C-----C-----	AF435967
1	1	-----G-----G-----A-----C-----	
1	1	-----G-----G-----A-----C-----	
1	1	-----G-----G-----A-----C-----	

## B PERV transfers in human APOBEC3G experiments

<u>N</u>	<u>Expt</u>	<u>Sequence</u>	<u>GB #</u>
17	1	GACCAGTATGGCCACAGCTGCGATAGCCTTCAGACATACGGGCCAACCACTGGCTACAGGATCAAGCTTCTTTGACAGGTAGGCAACAGGTCTCCTCCATGGTCCTAGGGTTTGGGTTAAAACTCCTCGGGCTACTCCCTTACGCTC	AY099323
8	2	-----G-----	AY099323
3	1	-----G-----	
1	1	-----A-----	

Jonsson *et al.*, Figure S4

