Dynamics of the Mammalian *APOBEC3* Locus and the Relationship Between Mammalian APOBEC3 and Lentiviral Vif Proteins

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Rebecca St. Claire LaRue

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor, Reuben S. Harris, Ph. D.

August 2010

Acknowledgements

I would like to acknowledge and thank my advisor, Reuben Harris, for inviting me in to his lab and enthusiastically supporting my comparative research interests. I would also like to thank Reuben for his support and accommodation in the last two years as I transitioned into being a new parent.

I would like to thank my family for being so supportive of my desire to pursue this degree. It would not have been possible without the many years of encouragement and support to follow my interests.

I would like to thank all of the Harris lab members past and present for making the lab a productive, supportive and fun environment. I would like to specifically acknowledge Stefan Jónsson for introducing me to the artiodactyl APOBEC3 research and continued collaboration, Joy Lengyel and Jessalyn Weaver for their help with experiments in Chapter 3, and Judd Hultquist for his superior editing skills for Chapters 1 and 4.

I would like to thank all our collaborators who contributed to my thesis research: Vala Andrésdóttir at the University of Iceland, Tim Smith and colleagues at the USDA, Isidro Hötzel at Genentech, Kevin Silverstein at the University of Minnesota Cancer Center, Matthew Lajoie, Denis Bertrand and Nadia El-Mabrouk at the University of Montreal and Scott Fahrenkrug at the University of Minnesota.

I would like to thank all of my committee members, Michael Murtuagh, Kent Reed, Perry Hackett and David Largaespada for advice and suggestions.

I would like to thank the Hendrickson, Bielinksy, Sobeck, and Livingston lab members for the lively Friday afternoon meetings.

I would like to thank the University of Minnesota's Institute for Molecular Virology and the Center of Genome Engineering.

I would like to thank the Comparative Molecular Bioscience graduate program for supporting me all these years: thanks to Mark Rutherford and Bruce Walcheck for their continued support of my research track and frequent recommendations, and a very special thanks to Lisa Hubinger for her excellent help in navigating the graduate program requirements.

I would like to acknowledge financial support from Spring Point Projects, and the Minnesota Agricultural Experimental Station and College of Veterinary Medicine.

Dedication

This dissertation is dedicated to my husband Christopher and my daughter Elinor.

Abstract

The mammalian immune system must be dynamic in the face of a diverse and ever-changing array of pathogens. Successful pathogens have evolved to overcome host immunity. One of the most successful pathogens in the last century is the human immunodeficiency virus (HIV), which is the etiological agent responsible for the global acquired immune deficiency syndrome (AIDS) epidemic. As a result, a massive amount of research is dedicated to determining how this retrovirus counteracts both innate and adaptive immune response in humans. Currently, one of the most studied host-pathogen interactions is between the cellular anti-retroviral APOBEC3 (A3) proteins and HIV viral infectivity factor (Vif) protein. A3 proteins are cytosine deaminases that primarily inhibit retroviruses and retrotransposons by mutating cytosines to uracils in retroviral DNA during reverse transcription. When HIV Vif is present, it can bind to certain A3 proteins and recruit cellular degradation complexes to target these anti-retroviral proteins for proteasomal degradation. HIV-like viruses (lentiviruses) infect other mammals and are the inspiration for studying A3 proteins in other mammals. The ultimate goal of this comparative study was to gain a better understanding of how the lentiviral Vif protein counteracts host A3 proteins.

The first component of this dissertation is dedicated to the process of determining the complete A3 protein repertoires of cow and sheep (representatives of the artiodactyl lineage), which are both infected with a lentiviruses. In order to achieve the total number of proteins of these artiodactyls, the *A3* genomic loci were sequenced and cDNAs were analyzed. It revealed that these particular mammals possess three *A3* genes that express

four proteins, which is significantly less than the seven human genes that encode seven proteins. These findings also helped to establish a foundation for understanding the evolutionary history of mammalian *A3* genes. The second component of this dissertation was to test if these artiodactyl A3 proteins were functional. Cow and sheep A3 proteins demonstrate intrinsic DNA cytosine deaminase activity and localize in cells similar to human A3 proteins. Furthermore, certain cow A3 proteins are capable of restricting HIV and are neutralized in the presence of cattle-specific lentiviral Vif (bovine immunodeficiency virus). The last component of this dissertation, a panel of conserved mammalian A3 proteins were tested to determine if each A3 protein interacts specifically with its species lentiviral Vif protein. It was shown that each representative host A3 protein is degraded in the presence of its species specific lentiviral Vif, suggesting a conserved interaction between mammalian A3 proteins and their species specific lentiviral Vif protein.

Table of Contents

List of Tables

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 1: Introduction

Overview

In order for a virus to successfully enter, replicate and reproduce in a cell it must overcome many innate-immune barriers. An example of a virus that has overcome these challenges (i.e. a successful virus) is the human immunodeficiency virus (HIV), the causative agent of acquired immunodeficiency syndrome (AIDS). HIV can enter, replicate and reproduce progeny in its primary host cell CD4+ T lymphocytes. However, if the viral infectivity factor (Vif) gene is disabled in the HIV genome, the virus is no longer able to infect its host target cells efficiently. This is because at least two human proteins – APOBEC3G (A3G) and APOBEC3F (A3F) – are efficiently able to restrict the replication Vif-defective viruses. The discovery of this host-pathogen protein-protein interaction has become the focus of intense investigation because of its therapeutic potential.

HIV is only one member of a large lentivirus family. Other lentiviruses can successfully infect other mammals such as cattle, sheep and cats. APOBEC3 (A3) proteins also exist in these mammals. This lead us to hypothesize that lentiviral Vif proteins neutralize host A3 proteins by a conserved mechanism similar to how HIV neutralizes the human A3G and A3F. The second chapter of this dissertation is dedicated to characterizing the complete A3-protein repertoire in cattle and sheep as well as learning more about the evolution and functionality of mammalian A3 genes. In the third chapter, the goal is to investigate if the cattle A3 proteins share a conserved anti-retroviral function with the human A3G and A3F proteins. A conserved, anti-retroviral

A3 protein in cattle and sheep, as well as humans, rhesus macaques and cats was tested for its ability to be neutralized by each species lentiviral Vif protein. The main conclusion from these studies is that the artiodactyl lineage contains fewer A3 genes than the primate lineage but express certain proteins that are capable of restricting HIV. Lastly, like human A3 proteins, A3 proteins of lentivirus-infected mammals are neutralized by their species specific lentiviral Vif.

Lentiviruses and the accessory protein Vif

Lentiviruses are a unique genus of complex retroviruses whose genomes are characterized by a number of additional genes beside the *gag*, *pol* and *env* genes. Some of these additional genes encode accessory proteins that are important for target cell infection but not always necessary for infection in experimental cell lines (reviewed in (162)). There are five mammalian lentivirus subgroups: (1) human immunodeficiency viruses type-1 and -2 (HIV-1, HIV-2, or collectively HIV); simian immunodeficiency viruses (SIV) in multiple species of primates; (2) feline immunodeficiency viruses (FIV) in domestic and wild cats; (3) bovine immunodeficiency viruses (BIV) in different species of cattle; (4) maedi-visna virus (MVV) in sheep and caprine arthritis encephalitis virus (CAEV) in goats; and (5) equine infectious anemia virus (EIAV) in horses (**Figure 1-1**)(173). Two of these additional genes, *rev* and *tat*, are present in all mammalian lentiviruses while other genes are present only in certain lentivirus subgroups (**Figure 1-1**). The functional importance of accessory proteins in the primate subgroup has been studied extensively but many of the accessory proteins of non-primate lentiviruses remain to be characterized. However, the viral infectivity gene (Vif) is present in four of the five

subgroups. Despite little overall sequence identity (< 30 %), lentiviral Vif proteins all possess a conserved S/TLQY/RLA motif in the C-terminal end that resembles a SOCs box (suppressor of cytokine signaling box)(**Figure 1-2 and Figure 1-4**) [reviewed in (65) and (124, 151)]. Mutating this motif in the *vif* gene prevents HIV from successfully infecting host target cells (primary CD4+ T-lymphocytes)(44)(45)(47)(129)(89, 93)(157). Similar observations were reported when this conserved motif was mutated or deleted in the *vif* gene in the lentivirus genomes of SIVmac, FIV, MVV and CAEV (34, 55, 81, 99, 170). These findings in different mammals provided evidence that the function of the Vif protein may be conserved in all Vif-encoding lentiviruses to successfully infect the host target cells.

Discovery of APOBEC3 proteins

The APOBEC (apolipoprotein B mRNA-editing catalytic polypeptide-like) protein family consists of polynucleotide cytosine deaminases that include APOBEC1, activation induced cytosine deaminase (AID), APOBEC2, APOBEC3 (A3) and APOBEC4. Members of this protein family contain either one or two zinc-coordinating domains (Z-domains) that share a conserved amino acid motif (HXE-X_{23-28}-PCX_{2-4}P). This motif is important for the enzymatic function of these proteins, the mutating of cytosines to uracils in single-stranded DNA and/or RNA. The first member to be identified was APOBEC1 and is the reason for the complicated acronym assigned to this protein family. Expressed primarily in the intestinal and liver tissue in humans, APOBEC1 deaminates a specific cytosine in the apolipoprotein B mRNA that generates a premature stop codon (168, 183). The result of this premature stop is a shorter form of apolipoprotein B, and

with the longer form of the protein, both play a biological role in lipid metabolism (5). The AID protein is essential for the diversification of immunoglobin genes. Expressed predominately in activated B cells, the ability of AID to mutate cytosines in the immunoglobin locus is essential for somatic hypermutation and class switch recombination (39, 117, 118, 132). The APOBEC2 and APOBEC4 proteins are members of this family solely based on possessing the conserved zinc-coordinating motif, however, the biological function of these proteins remains unclear. APOBEC2 expression is detected in cardiac and skeletal tissues but it does not appear to deaminate cytosines in DNA or RNA (58, 96, 107, 133). The APOBEC4 gene was identified by computational analysis of database sequence and assigned to the APOBEC family based on the conserved zinc-coordinating motif (139).

The A3 proteins are the most notable members of the APOBEC family and are the focus of the studies presented in this dissertation. A3G was the first A3 protein member discovered by three independent experimental studies (58, 66, 153). In the first study, Jarmuz et al. (66) identified a locus on human chromosome 22 of seven APOBEC1-like genes arranged in tandem (A3A, A3B, A3C, A3DE, A3F, A3G and A3H) (**Figure 1-2**). These A3 members share either one (single domain A3A, A3C, A3H) or two (double domain A3B, A3DE, A3F, A3G) conserved zinc-binding domains (HXE-X_{23-28}-$PCX_{2-4}P$). In the second study by Sheehy et al.(153), A3G was identified as a cellular factor that inhibits infection of Vif-deficient HIV. This finding, as well as the experimental platform for this discovery, provided an explanation for why some human T-cell lines were permissive and some non-permissive to HIVΔVif infection. In the third study, Harris et al.(58) investigated the DNA-editing properties of the human mRNA editing APOBEC1

protein (prototypic A3 family member) as well as A3G and A3C in comparison to another related DNA-editing family member, AID. Using an *E. coli* experimental system to investigate intrinsic mutator activity, like AID and APOBEC1 proteins, A3G and A3C were able to deaminate cytosines to uracils in genomic DNA there by causing high frequencies of cytosine/guanine to thymine/adenine (C/G to T/A) mutation. Since these original studies, A3 proteins have been found to play an important role in protecting the integrity of a host genome by restricting the infection of exogenous pathogens, the persistence of foreign DNA, and the movement of endogenous retroelements (reviewed in (63, 160)).

Human APOBEC3 proteins and HIV restriction

Much of the research on the A3 proteins has been directed at their ability to inhibit retroviruses, mainly the ability of A3G to restrict Vif-deficient HIV. Studies by different groups have shown that retroviral restriction is largely dependent on A3G's ability to deaminate cytosines to uracils in single-stranded DNA (58, 90, 102, 190). In the absence of Vif, this 'hypermutation' phenotype was found in retroviral cDNA produced in the presence of A3G. This ability to hypermutate cDNA was mapped to the C-terminal zinc-binding domain of A3G through mutational analysis of the conserved zinc-binding residues (**Figure 1-3**) (52, 120, 121). A model has been proposed for the involvement of the A3G protein in the HIV life cycle (**Figure 1-4**) reviewed in (57). A3G gains access to the HIV particles during the packaging process in an infected cell. When the A3G-containing virus particle infects a new cell, A3G is able to deaminate cytosines to uracils after synthesis of the retroviral DNA minus strand from the RNA genome. During plus-

strand synthesis, the viral reverse transcriptase recognizes the uracil as a thymine, and an adenine is incorporated into the newly synthesized complementary strand of DNA. At this point in the HIV life cycle, either the newly synthesized retroviral provirus is degraded or it is integrated into the host's genome where it displays the hypermutation phenotype (C/G to T/A mutations). The accessory protein Vif is able to inhibit A3G by serving as an adaptor protein that links it to the CULLIN-5 based E3 ubiquitin ligase complex for polyubiquitination and subsequent degradation by the cellular proteasome before it is incorporated into the budding viral particle. (**Figure 1-4**)(187). While this model of HIV restriction is based mostly on studies with A3G, another human A3 (A3F) has been shown to be equally effective in restricting HIVΔVif and is thought to work by a similar mechanism (10, 97, 179, 194). Many studies have reported that the remaining A3 members (A3A, A3B, A3C, A3DE, A3H) can restrict HIVΔVif to varying degrees in single cycle HIV infectivity assays, but an almost equally large number of studies have been negative [reviewed in (3)].

Human APOBEC3 proteins and HIV Vif

Investigating the details of the A3G and A3F-HIV Vif protein interactions has been a main goal in the hope of developing effective therapeutics that could interrupt this interaction and allow host A3 proteins to restrict HIV. Initial studies utilized other primates A3Gs and SIV Vif proteins to identify the Vif sensitive region of A3G. HIV Vif and SIVagm Vif specifically neutralize their host A3G but not the other host's A3G (11, 103, 104, 147, 182). By swapping the residues at position 128 in both human (aspartic acid) and African green monkey (lysine) A3Gs, the proteins became sensitive to the other

species' Vif. Numerous studies followed, utilizing deletion and mutagenesis techniques to reconfirm that D128 was responsible for the sensitivity to Vif, as well as other nearby residues located in the NTD of A3G [reviewed in (156)] (**Figure 1-3**). Similar techniques were used to identify a patch of Vif-sensitive residues in the CTD of A3F [(142) and reviewed in (156)]. With the exception of A3A and A3B, the remaining A3 members, A3C, A3DE and A3H, have reported sensitivity to HIV Vif [reviewed by (3)]. Not all of these A3 proteins are able to restrict HIV so it is possible that a conserved Vif-interacting structure exists in all A3 proteins. The relatively recent high-resolution structures of the A3G CTD could help clarify a conserved Vif-interacting region by revealing the location of sensitive residues (20)(62)(43)(54)(152).

The S/TLQY/RLA motif in HIV Vif, conserved in all lentiviral Vif proteins, was discovered to be essential for binding of the ELONGINB/C heterodimer that is part of the CULLIN5-based E3 ubiquitin ligase complex. As noted above, HIV Vif serves as an adaptor bridge to this complex so that A3G can be polyubiquitinylated and marked for degradation by the cellular 26S proteasome (108, 187, 188). Many species of primate SIV Vifs also share a zinc-binding domain ($Hx_5Cx_2Cx_3H$) necessary for the binding of the CULLIN 5 protein (**Figure 1-3 and 1-5**)(181).

Numerous studies show that the A3-interacting region of the HIV Vif protein is located in the NTD (**Figure 1-3 and Figure 1-5**) (105)(109)(141)(147)(155)(169)(184)(193). Utilizing site-directed mutagenesis, biochemical changes were made to alter charge, and patches of residues were found to be specific to either A3G (109) (141)(184)(193) or A3F (60)(109)(141)(147)(155)(169) (184) (193), or were required for both A3G and A3F (18, 31, 60, 131, 141, 155, 184)

(**Figure 1-5).** It has also been reported recently, that other A3 family members A3C and A3DE bind to regions on Vif specific to A3F (193). To date, the data collected show that each A3 interacts with a different region of Vif but some of these regions overlap. A complete, high-resolution structure of Vif or co-structure of Vif and A3G/F could help visualize the topology of these multiply A3-interacting regions that are involved in this protein-protein interaction.

A3 proteins in other mammals

The first *A3* loci to be completely annotated were the human and mouse, and they revealed a large difference in the number of *A3* genes (**Figure 1-2**). As mentioned above, there are seven human *A3* genes on chromosome 22, whereas in the mouse there is only one *A3* gene (66)(177). In a study by Conticello et al. (26), phylogenetic analysis of each zinc-coordinating (Z) domain in the human A3 proteins (11 total domains) and A3 mouse protein (2 domains) showed three distinct groups referred as Z1, Z2 and Z3 (nomenclature reflects (87), Appendix 1) (**Figure 1-2**). The mouse A3 protein contains only the Z2 and Z3 of these different Z-type domains and in an organization (Z2-Z3) not present in human A3 proteins (**Figure 1-2**). Soon after this study, phylogenetic analysis of other mammalian A3 Z-domains, mined from partially sequenced genomes and cDNA in nucleotide and protein databases, confirmed the presence of Z1-, Z2- and Z3-type A3 domains in cattle and the Z2- and Z3-type A3 domains in pigs and cats. However, it was still uncertain if this Z1-type, or more Z2- and Z3-type domains existed in these mammals because of incomplete sequence of the A3 genomic loci and the possibility that

A3 genes are not transcribed in the tissues from which cDNA libraries were made and submitted to the database.

Shortly following the phylogenetic analysis of the available A3 sequences in non-primate mammals, a comprehensive study was performed on the conserved A3Z2-Z3 protein in pig, cow and sheep (68). These non-primate A3 proteins were shown to be functional cytosine deaminases that can restrict HIV and murine leukemia virus (MLV). A similar study showed that the cat A3Z2-Z3 and a single Z3-type A3 can restrict HIV and deaminate the viral genome (116). So the presence of functional A3-like genes and proteins and studies showing that Vif-deficient lentiviruses cannot efficiently infect their host-target cells [discussed above(34, 55, 81, 99, 170)], provided evidence that the human A3G/A3F-HIV Vif protein interaction may be conserved between lentiviral Vif and host A3 proteins of other mammals.

All in one year, the complete A3 loci of the cat (115), cow and sheep [Chapter 2 (88)] were reported. Supporting functional studies confirmed that these mammals have genes that express functional cytosine deaminases (Chapter 2). This information created a foundation to address the hypothesis that species specific lentiviruses are able to evade the repertoire of host A3 proteins using a Vif-like mechanism similar to that which enables HIV to evade A3G and A3F. In order to address this hypothesis, several important questions needed to be answered. First, it was essential to define the complete A3 repertoire of mammals infected with Vif-containing lentiviruses (Chapter 2 and other studies). Second, the A3 proteins of these mammals needed to be tested for their ability to restrict virus and be neutralized by their species lentiviral Vif (Chapter 3 and other studies). The ultimate goal of these studies was to gain a more robust understanding of

the human A3/HIV Vif protein interaction by investigating how other species proteins have evolved to counteract each other.

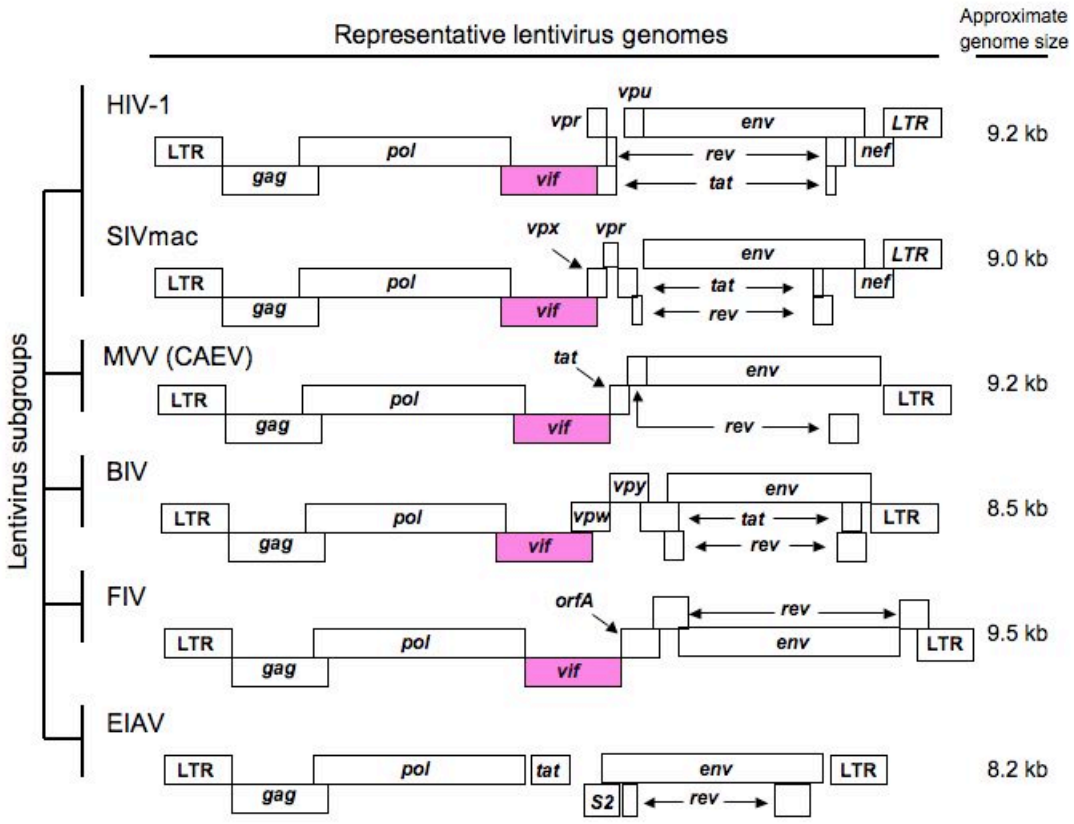


Figure 1-1. Comparison of lentiviral genome organization. The five subgroups of lentiviruses [based on (173)] are represented with species from their particular lineage: HIV and SIVmac (primates), BIV (bovid), MVV (ovid/caprine)(CAEV genome not shown), FIV (feline) and EIAV (equine) [adapted from(24)(93)(124)(151)]. The Vif gene is highlighted in purple. Genomes are drawn approximately to scale and are based on these lentivirus strains: HIV_{IIIB} (gb EU541617), SIV_{mac239} (gb AY588946), BIV_{BIM127} (gb M32690), MVV Icelandic strain 1514 (gb M60610), FIV_{NSCU} (gb M25381), and EIAV (gb AF247394).

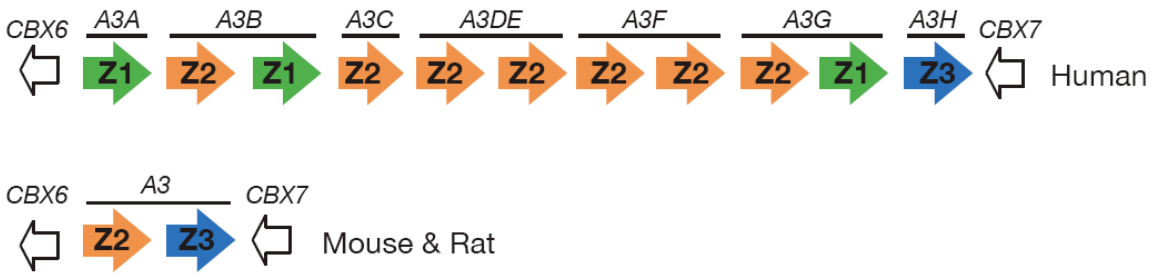


Figure 1-2. A schematic of the *A3* locus in humans and mice (adapted from (87)). Arrows represent an A3 domain containing the conserved zinc-coordinating domains. The phylogenetically distinct groupings of the conserved zinc-coordinating domains are indicated by labeling but also color (Z1-green, Z2-orange and Z3-blue) [phylogenetic analysis based on (26) and nomenclature based on (87)]. The lines on top indicate if the protein expresses one or two zinc-binding domains. The open arrows indicate the *CBX6* and *CBX7* genes that flank both of these mammals *A3* loci.

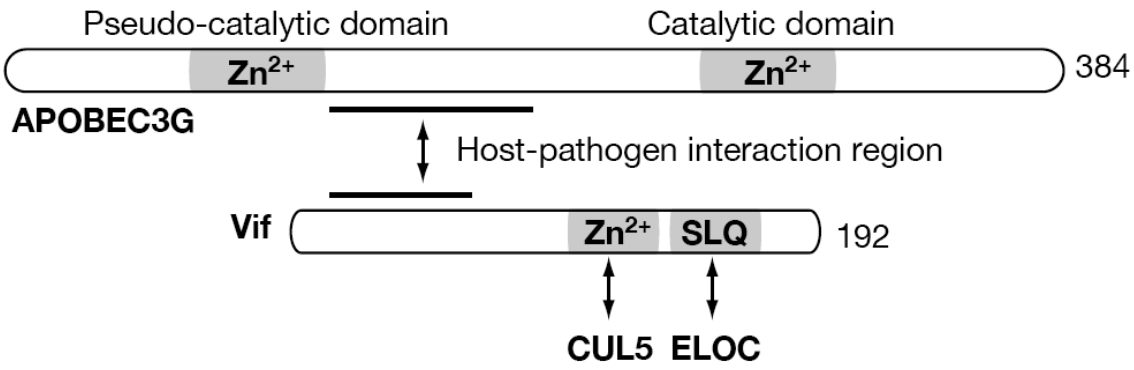


Figure 1-3. A schematic of human A3G and HIV Vif indicating some relevant motifs. CUL5= Cullen-5 protein, ELOC= ELONGIN C protein. Figure adapted from (59).

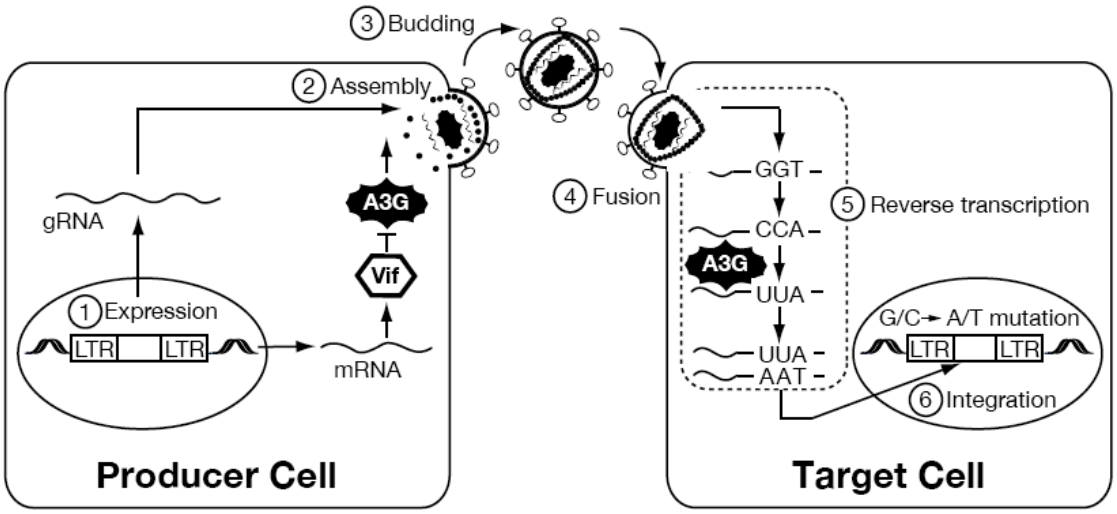


Figure 1-4. A current model for HIV restriction by human A3G depicted within the context of a simplified 6-step viral life cycle. Producer cell A3G encapsidates through Gag ribonucleoprotein interaction (not shown) travels with nascent viral particles until a new target cell is infected, and deaminates viral cDNA during reverse transcription. The dotted line surrounding the reverse transcription and DNA deamination steps indicates that these are generally thought to occur within the confines of a semi-porous capsid shell. Figure adapted from (59).

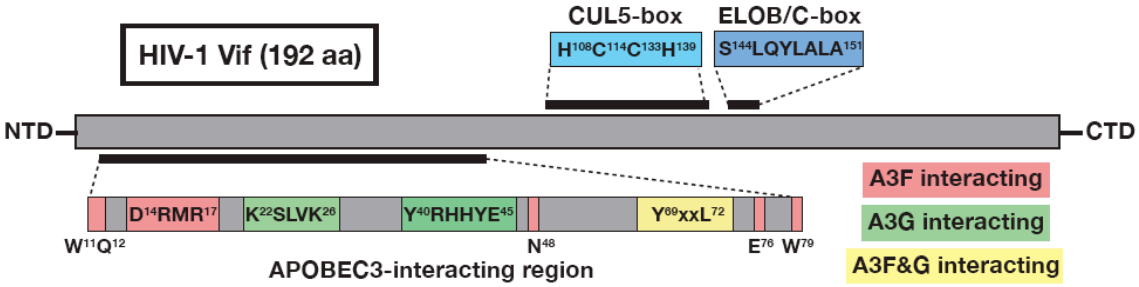


Figure 1-5. A schematic of functional motifs in HIV Vif. Elongin BC (ELOB/C) and CULLIN 5 (CUL5) motifs are based on primate lentiviral Vif sequences. A3-interacting regions are coded by color (A3F = green, A3G = red, A3F and A3G = yellow). All references cited in text.

CHAPTER 2

The Artiodactyl *APOBEC3* Innate Immune Repertoire Shows Evidence for a Multi-functional Domain Organization that Existed in the Ancestor of Placental Mammals

Reprinted with permission from: R.S. LaRue, S.R. Jónsson, K.A.T. Silverstein, M. Lajoie, D. Bertrand, N. El-Mabrouk, I. Hötzel, V. Andrésdóttir, T.P.L. Smith, and R.S. Harris (2008) BMC Mol. Biol.

Authors' contributions

R. S. LaRue and R.S. Harris designed the studies, performed experiments, analyzed data and wrote the manuscript. S.R. Jónsson and V. Andrésdóttir helped analyze the artiodactyl A3 genes and proteins, T.P.L. Smith provided library samples and generated genomic DNA sequences, I. Hötzel contributed cattle A3 gene sequences and functional data, K.A.T. Silverstein assisted with phylogenetic and computational studies, and M. Lajoie, D. Bertrand and N. El-Mabrouk generated the model for A3 evolution. All authors contributed to editing the manuscript.

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: cattle, pigs and sheep. 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. This dynamic recombination-filled history is consistent with a model in which most of these events were selected by ancient pathogenic retrovirus infections during primatification.

Introduction

Mammalian APOBEC3 (A3) proteins have the capacity to potently inhibit the replication of a diverse set of reverse-transcribing mobile genetic elements (8)(22)(28)(50)(101). Susceptible exogenous retroelements include lentiviruses (HIV-1, HIV-2, several strains of SIV and FIV), alpharetroviruses (RSV), betaretroviruses (MPMV), gammaretroviruses (MLV), foamy viruses and the hepadnavirus HBV [*e.g.*, (35)(56)(70)(102)(127)(163)(171)(178)(190)]. 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 (6)(37)(38)(67)(69)(71)(91)(92)(149). In addition, some A3 proteins can also inhibit L1 and its obligate parasite Alu, retrotransposons that replicate by integration-primed reverse transcription (19)(76)(172)(113)(14)(122)(158). 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 HTLV and MLV appear to simply employ an avoidance mechanism [*e.g.*, (2)

(33)(143)(187)}]. 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_1-E-x_{25-31}-C-x_{2-4}-C [x indicates a non-conserved position; (26, 57)]. The A3 Z domains can be grouped into one of three distinct phylogenetic clusters - Z1, Z2 or Z3. (**Figures 2-1 & 2-8**). The Z-based classification system, proposed originally by Conticello and coworkers (26), was revised recently through a collaborative effort (87). From hereon, the new A3 nomenclature system will be used. Z1 and Z2 proteins have a SW-S/T-C-x_{2-4}-C motif, whereas Z3 proteins have a TW-S-C-x_2-C motif. Z1 and Z2 proteins can be further distinguished by H-x_1-E-x_5-X-V/I and H-x_1-E-x_5-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.*, (52, 120, 121, 126)].

We previously reported a double-domain *A3Z2-Z3* gene *(*formally *A3F)* from the artiodactyls, sheep (*Ovis aries*), cattle (*Bos taurus*) and pigs (*Sus scrofa*) (68). 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; (68)]. 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*. Two BACS were identified, one spanning the

entire *CBX6* to *CBX7* region and the other including only *A3Z2, A3Z3* and *CBX7*. These BACS were 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 2-2A**). 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 2-2A**; Btau_4.0 http://www.hgsc.bcm.tmc.edu/projects/bovine/). 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 (MYA) (9, 83, 138).

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 (9, 83), 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 2-2A**).

The cattle, sheep and pig *A3* locus genomic sequences were compared using dotplot analyses (**Figures 2-2B & 2-9**). 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 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 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 (9). A pan-species, *A3Z1* PCR primer set was developed and used in these experiments. In contrast to human, 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 2-2C**). 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 2-1C & 2-10**). 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 (9, 83) 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 (9, 83), 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 (68). 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 2-3**). This truncated transcript was readily amplified from sheep and cattle PBMCs and represented by existing EST sequences (**Figure 2-10** 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 2-3 & 2-10**). 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 2-3 & 2-10**). 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 (68).

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 (**Figure 2-10** 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 upsteam of *A3Z2* (68). A comparison of cDNA and genomic sequences

revealed the presence of an exon in this location (*A3Z3* exon 1 in **Figure 2-3**). 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 2-3**).

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* (**Figures 2-3 & 2-11**). 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.*, (1, 15, 70, 130, 167)]. These putative ISREs significant similarity to functional elements in known interferon-inducible genes *ISG54* and *ISG15* (74, 94, 136). 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 [(114, 115); LaRue & Harris, data not shown].

Together with our previous data on the double domain A3 protein of these artiodactyl species, A3Z2-Z3 (68), 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 modularity was reported recently for the cat A3 locus, where two single domain *A3* genes combined to produce a functional double-domain A3 protein (115). 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 – A3Z2-Z3, A3Z2, A3Z3 and A3Z1 – 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.*, (19, 30, 58, 68, 97, 104, 115, 126, 186)]. For instance, we showed that the artiodactyl A3Z2-Z3 proteins could catalyze the deamination of *E.coli* DNA and retroviral cDNA (68). 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.*, (52, 64, 68, 120, 121, 158)]. 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 (13). 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^R) mutations in *E. coli* provide a quantitative measure of the

intrinsic A3 protein DNA cytosine deaminase activity [*e.g.*, (52, 58, 97, 121)]. In contrast to full-length cow A3Z2-Z3, which triggered a modest 2-fold increase in the median Rif^R mutation frequency over the vector control, non-induced levels of cow A3Z2 caused a large 50-fold increase (**Figure 2-4A**). 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^R mutation frequency (**Figure 2-12** 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^R mutation frequency (**Figure 2-4B**). This result contrasted with the related Z3 protein of humans, A3H, which appeared inactive in this assay (**Figure 2-4B & 2-10**). 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 (29, 115, 125, 126). Thus, our intended human A3H control appears to be the exception rather than the rule and that the single-domain 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 2-4C** 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 (13, 19, 21, 158). 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^R 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 2-4C**). To summarize this section, all four of the A3 proteins of artiodactyls demonstrated intrinsic DNA cytosine deaminase activity.

A3Z2-Z3, A3Z1, A3Z2 and A3Z3 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 2-5** 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 2-5**, 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 2-5**, row 1; (68, 69)]. 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

(115, 126, 146, 191). 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 event neutral or negative). This possibility was tested first by using the PAML free ratio model to calculate the number of mutations that resulted in amino acid replacements over the number that were silent between pairs of artiodactyl species. This ratio of replacement (dN) to silent (dS) mutations yields an omega (ω) 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.

A combined phylogeny was generated for each distinct *A3* Z domain and its inferred ancestral sequences (**Figure 2-13**). Both the artiodactyl *A3Z1* and the *A3Z2* genes appeared to be under a weak negative selection pressure, with ω values uniformly below one (**Figure 2-13A & B**). 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 (**Figure 2-13C**). However, a comparison of the inferred ancestral ruminant sequence with the inferred porcine sequence yielded a ω value of 1.5, suggesting that the ancestor(s) of modern day artiodactyls may have experienced intermittent positive selection (**Figure 2-13C**). These data clearly contrasted with primate *A3H* and ancestral primate *A3Z3* sequence comparisons, which showed evidence for very strong positive selection [originally reported by (126); re-calculated here with a representative clade shown in **Figure 2-13C**]. 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, a more focussed examination of artiodactyl *A3* Z domain variation using PAML NsSites revealed some evidence for positive selection (**Table 2-1**; see **Methods** for procedural details and **Figure 2-10** for sequence information). A robust analysis was ensured with the use of several distinct positive selection models (M2 and M8 and two codon frequency models F61 and F3x4). 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 (**Tables 2-1 & Figure 2-13**). The PAML NsSites program is thought to be more robust than the free-ratio model used above. 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 2-6**; also see **Figures 2-1 & 2-8**). Molecular phylogenetic data helped us infer that such a common ancestor existed approximately 100-115 MYA (9, 83). 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 2-6** & **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 2-6**; (48, 126, 174) and our unpublished data]. Thus, since the human and macaque lineages diverged approximately 25 MYA (9, 48, 134), 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 2-6**).

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 (42, 46)]. 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 (26). 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.

An 8-event model for human *A3* Z domain history is shown in **Figure 2-7** (see **Figure 2-14** 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-$(Z2)_3$-Z1-Z3] was a key intermediate in the evolution of the primate *A3* locus (event 4 product in **Figure 2-7**).

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 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 **Figures 2-7 & 2-14**). 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 2-6**).

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 (26, 57, 66, 68, 115)]. 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 repertoire, 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 2-6**). 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 phylogentic branch set. These data supported 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). 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 2-2C**; (110, 175)]. 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 [(126, 127, 146, 191) and this study, **Table 2-1**]. 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 2-6**). 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 *A3*s 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 2-5**). An *A3Z1* deletion was shown here for pigs, inferred here for cats and mice/rats, and demonstrated recently for some human populations [**Figure 2-6** and (75)].

Finally, a major question is what selective pressure(s) drove the largely primate-specific *A3* expansion from an ancestral Z1-Z2-Z3 repertoire to the present day human Z1-Z2-Z1-$(Z2)_6$-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-$(Z2)_6$-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 the 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 th 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 ancient retroviral infection.

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; (40)]. *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 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) (16). Repetitive sequences were identified using RepeatMasker (www.repeatmasker.org).

A3 exons were identified by directly comparing the genomic DNA sequences with cDNA, EST and RACE sequences [below, **Figures 2-10 & 2-15** and (68)]. Predicted ISREs were identified and compared using the TransFac and Biobase databases through the softberry NSITE portal (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* Landrance/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 **Figure 2-15**.

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 (58, 68). 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 (41) 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 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 (68, 159). 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 [(112); 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^R DNA deamination assays

Cytosine deaminase activity of the artiodactyl A3 protein variants was measured by quantifying the accumulation of Rif^R mutants in *ung*-deficient *E. coli* [*e.g.*, (58, 68)]. 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 24h in Labtek chambered coverglasses (Nunc), transfected with 200ng of the pEGFP-N3

based constructs and, after an additional 24h visualized on a Zeiss Axiovert 200 microscope at 400x magnification. HsA3F, HsA3C, HsA3H, HsA3A, BtA3Z2-Z3 and SsA3Z2-Z3 fusion constructs were previously reported (68, 69, 159).

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 (80). T_coffee version 5.31 was used for multiple sequence alignments (123). PAL2NAL software was used to convert amino acid sequences to nucleotides (164). JalView was used to remove insertions/deletions (23). The dnaml program within the Phylip software package was used to generate a phylogenetic tree [(137); an identical tree was obtained with MrBayes version 3.1 (140), except branch lengths differed slightly]. Clustal W version 1.83.1 was also used for some individual domain comparisons (86).

Free ratio model positive selection studies were based on a phylogenetic tree generated through Bayesian inference using MrBayes version 3.1 (140). Each tree was run for 250,000 generations with a burnin of 62,500 and standard default parameters. The PAML codeml program (185) was used to generate dN/dS ratios (ω values) for phylogenetic tree branches. ω values from the free ratio model using the F3x4 algorithm are shown in **Figure 2-13** (values from the F1x61 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 2-1**). 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 **Figure 2-10**). 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-$(Z2)_6$-Z1-Z3. The considered duplications are 'multiple tandem duplications' resulting from unequal crossing over (42). 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 (7, 46, 84, 192), 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 2-7 & 2-14**).

To infer the most recent evolutionary events (events 5 to 8 in **Figure 2-7 & 2-14**), 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 (www.repeatmasker.org). A dot plot of this sequence with itself was obtained using Gepard (82) to identify pairs of regions with very high similarities. The three most significant were extracted and further aligned using Blastz (150) 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 **Figure 2-10** for a full list of GenBank accession numbers).

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.

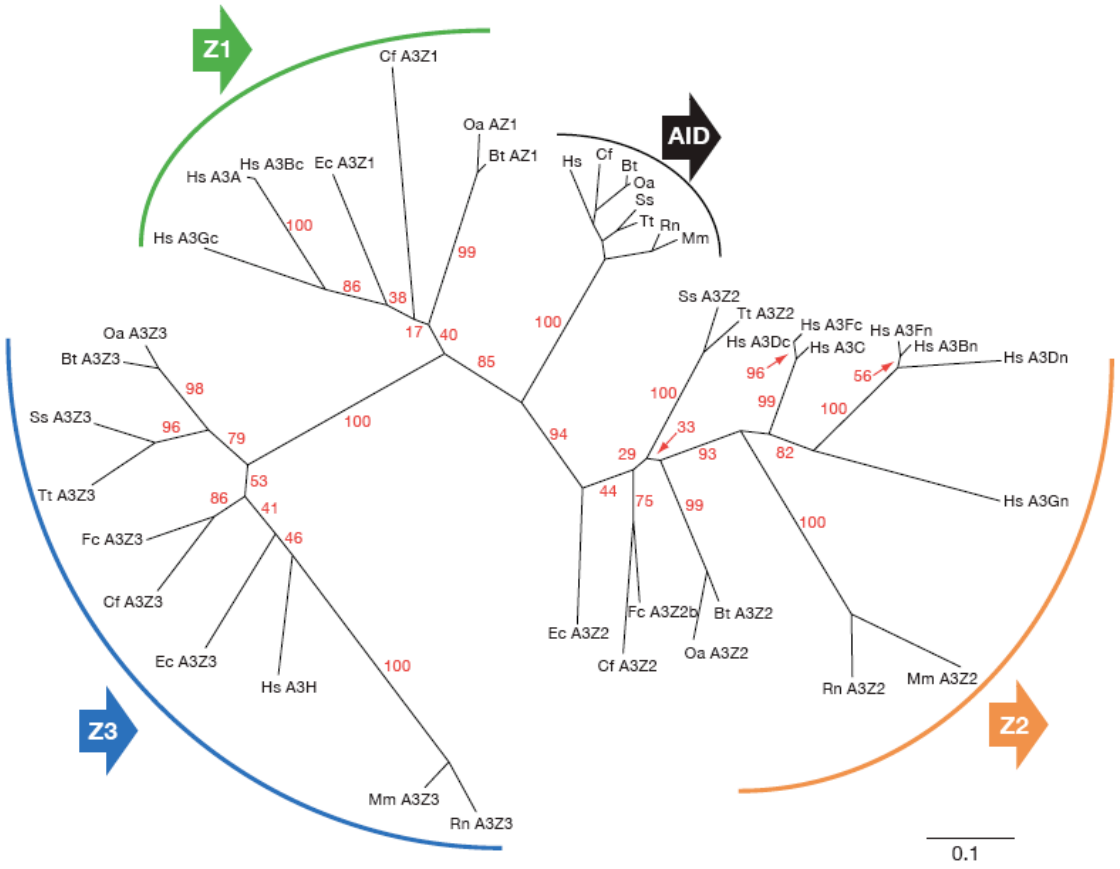


Figure 2-1. The mammalian A3 Z domains form three distinct phylogenetic groups. Bootstrap values are indicated in red. The scale bar represents 0.1 nucleotide changes per codon. See the Methods for details. Abbreviations for mammals: Hs=human, Bt=cow, Oa=sheep, Ss=pig, Tt=peccary, Ec=horse, Cf=dog, Fc=cat, Mm=mouse and Rn=rat. Other abbreviations: n=amino terminal domain and c=carboxy terminal domain.

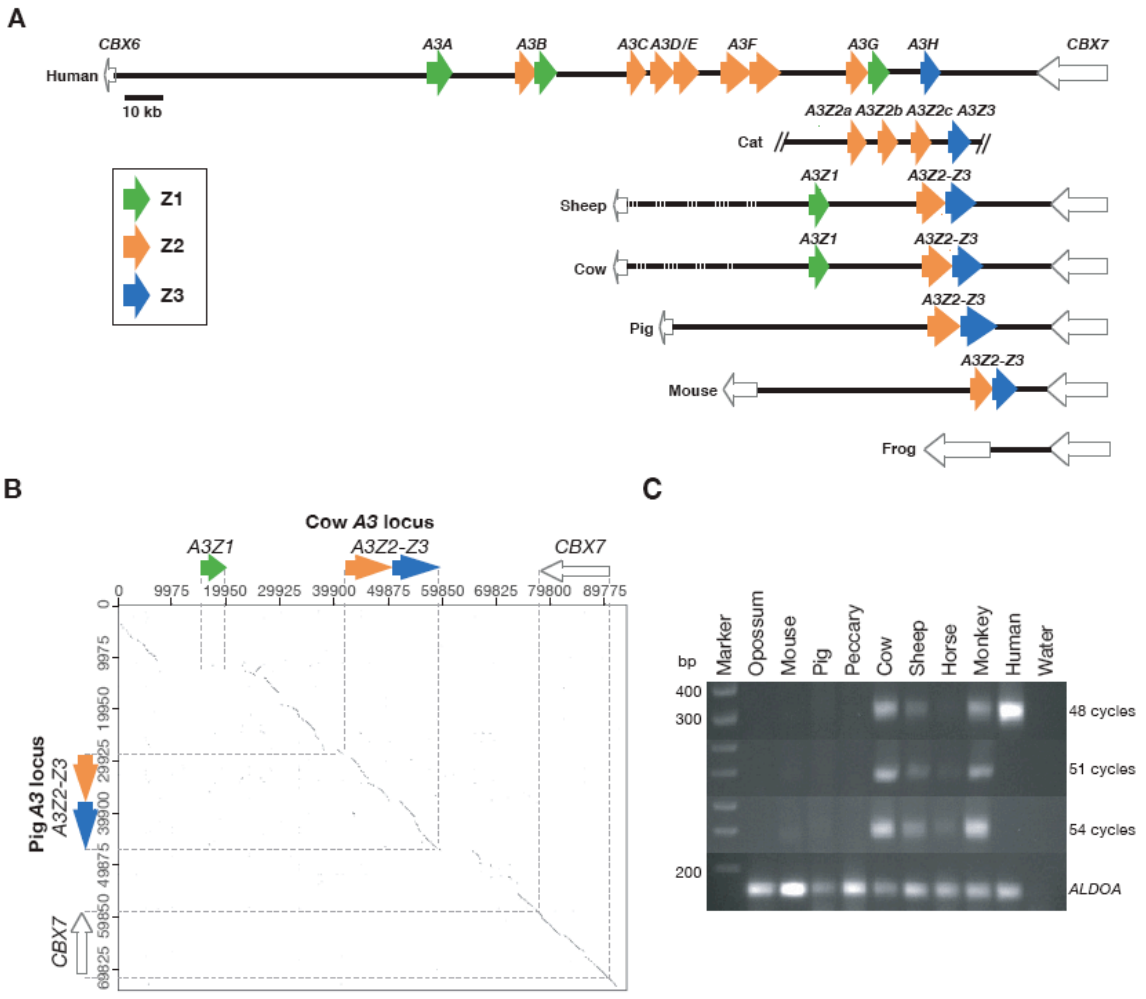


Figure 2-2. The *A3* genomic repertoire of sheep, cattle and pigs. (*A*) An illustration of the *A3* genes of the indicated mammals. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. The conserved flanking genes *CBX6* and *CBX7* are shown and the scale is indicated. Solid lines 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).

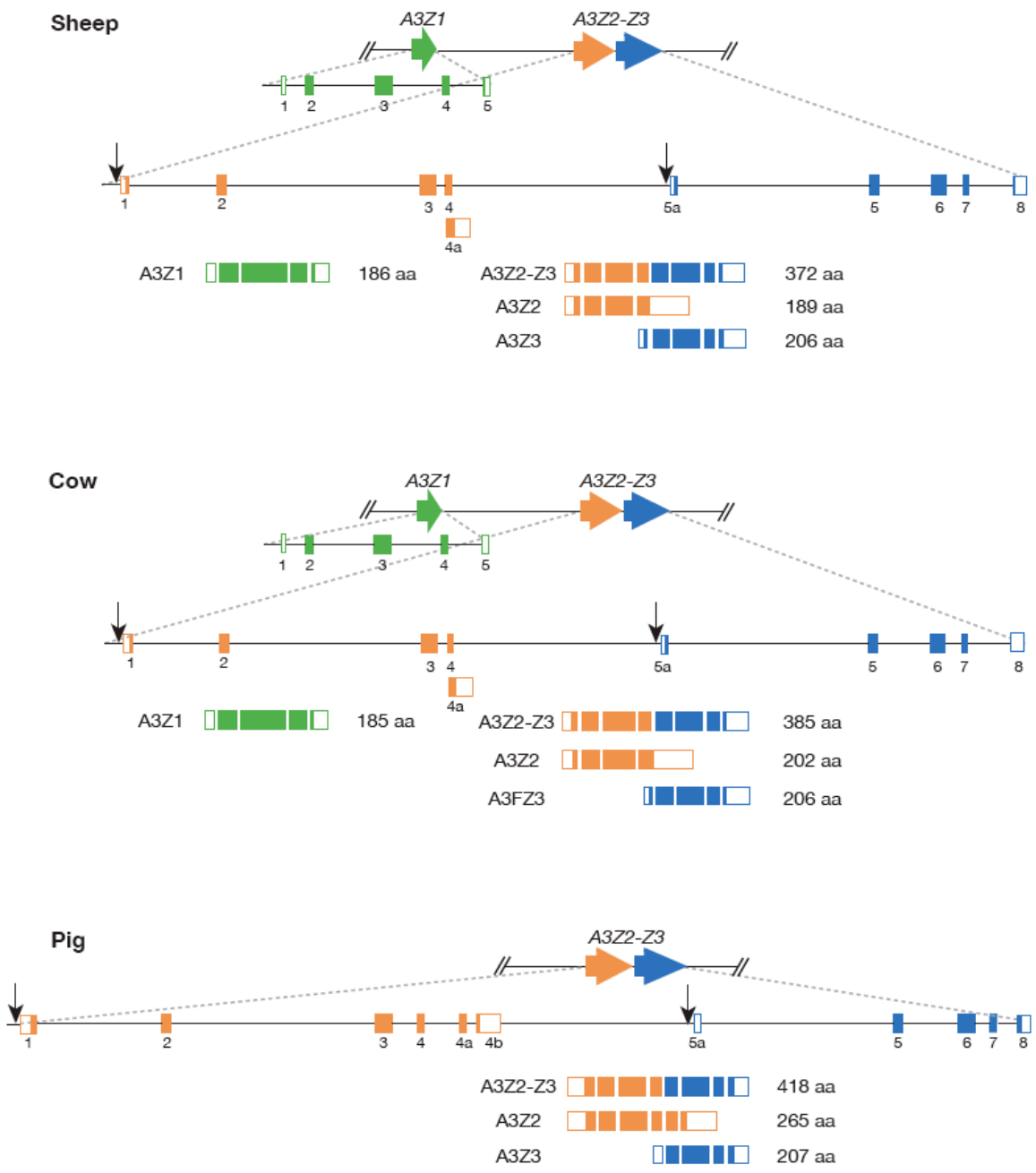


Figure 2-3. The coding potential of the sheep, cow and pig *A3* genes. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. The exons are shown below the gene schematics with coding regions represented by filled boxes and untranslated regions by open boxes. The gene schematics and exon blow-ups are drawn to scale. Arrows

indicate approximate positions of predicted ISREs (**Figure 2-11**). See the main text and the **Methods** for details.

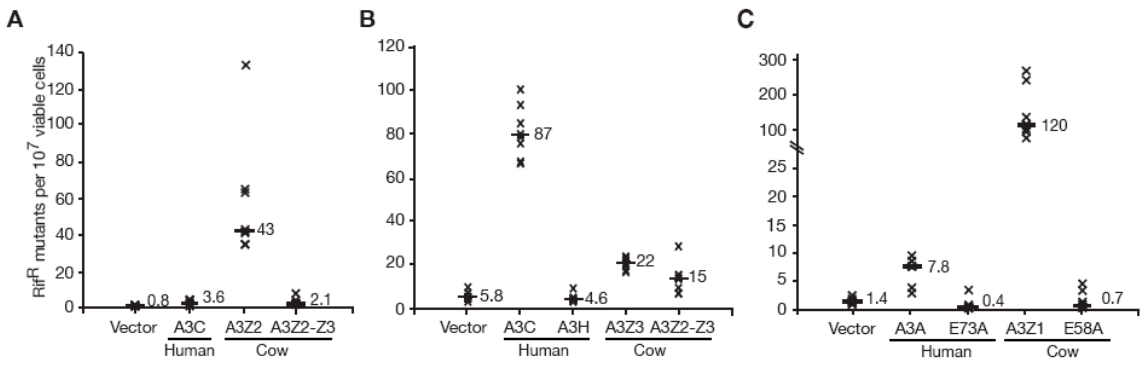


Figure 2-4. The artiodactyl A3 proteins catalyze DNA cytosine deamination. (*A*) Cow A3Z2 triggers a strong mutator phenotype in *E. coli*. Rif^R mutation frequency of 4-6 independent bacteria cultures expressing basal levels of indicated A3 proteins (each X represents data from a single culture). To 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).

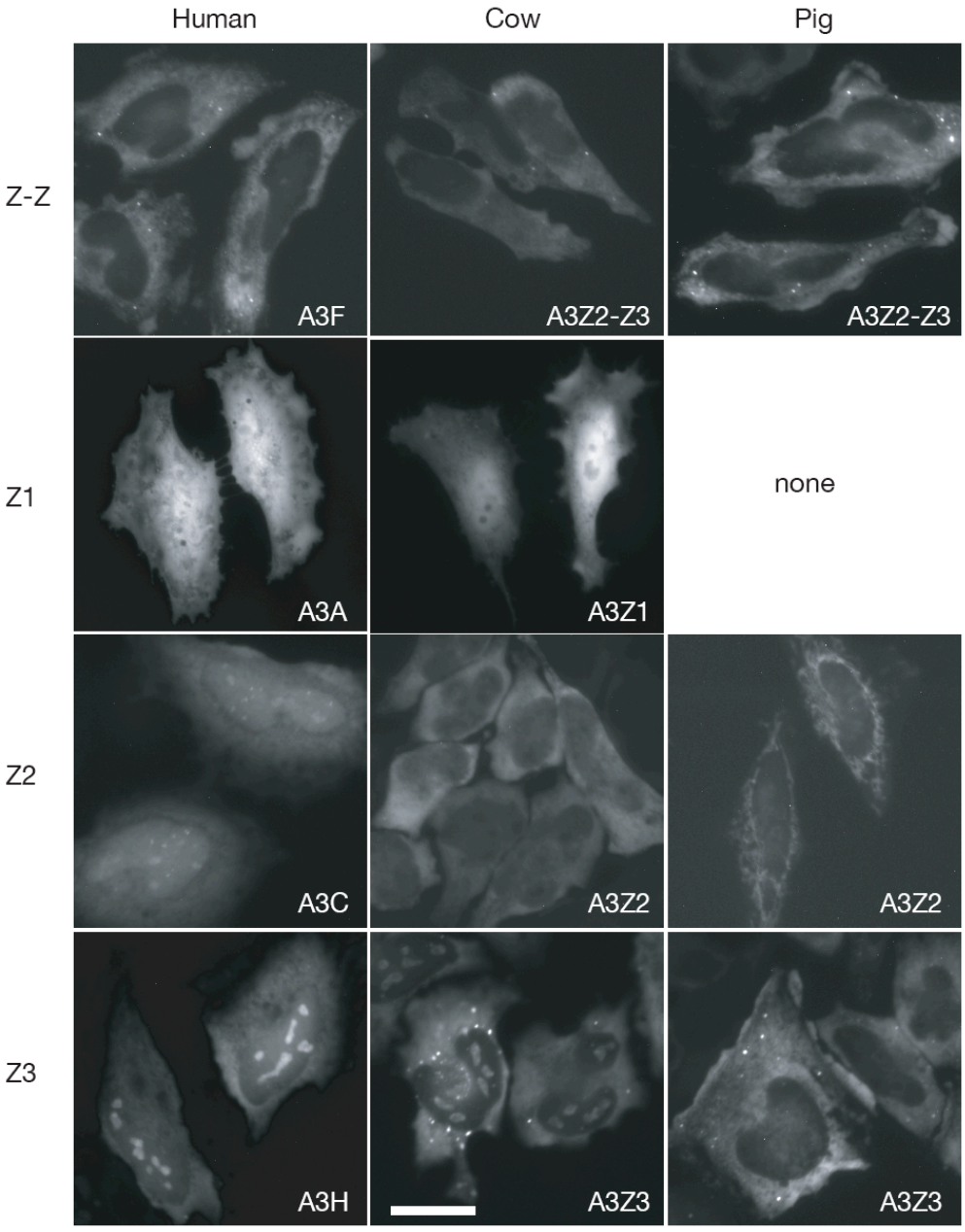


Figure 2-5. The subcellular distribution of cow and pig A3 proteins in comparison to human A3 proteins with similar Z domains. Representative images of live HeLa cells expressing the indicated A3-GFP fusion proteins are shown. The scale bar represents 25 μm.

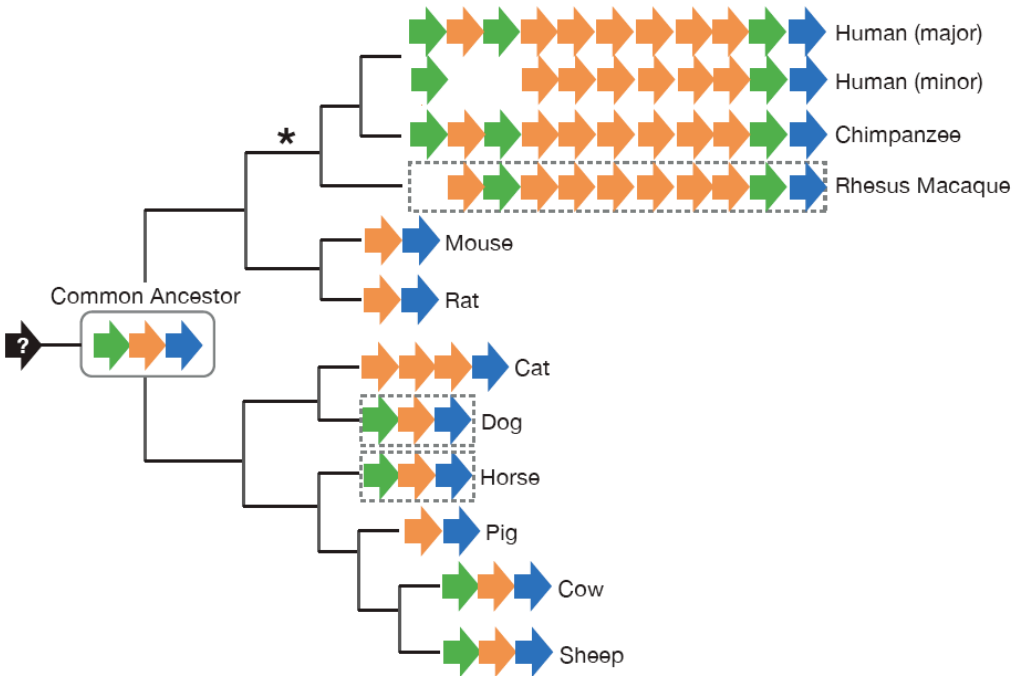


Figure 2-6. The distribution of *A3* Z domains in mammals. The common ancestor of the indicated placental mammals was inferred to have a Z1-Z2-Z3 *A3* gene repertoire. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. A question mark specifies the original *AID*-like ancestor. An asterisk indicates the likely period in which the bulk of the primate *A3* gene expansion occurred (see main text, **Figure 2-7** and **Figure 2-14**). 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 (**Figure 2-10**).

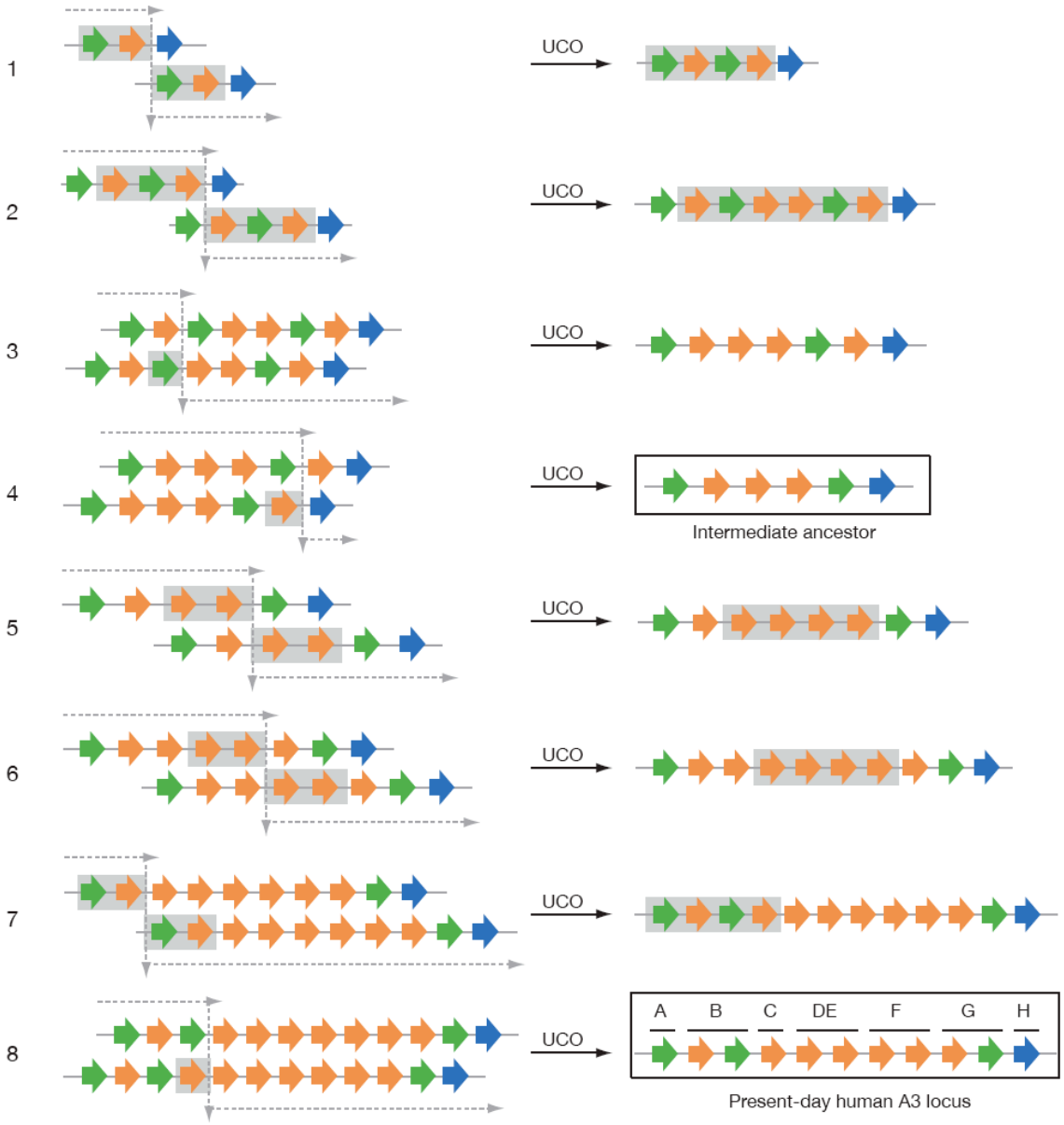


Figure 2-7. An 8-event model for the duplication and deletion history of the human *A3* repertoire. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. Five duplication and three deletion events were predicted to transform the ancestral locus into the present-day human *A3* repertoire. 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 2-1**, 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 **Figure 2-14** and details can be found in the main text and **Methods**.

Table 2-1. Evidence for positive selection in the artiodactyl Z domains.

Z domain[a]	Codon frequency model[b]	Comparison of null and positive selection models[c]	Significance	Tree length[d]	dN/dS (%)[e]
Z1	F61	M1-M2	p=0.01	2.6	2.5 (50)
		M7-M8	p=0.01	2.6	2.5 (50)
	F3x4	M1-M2	p=0.04	3.9	3.9 (28)
		M7-M8	p=0.02	3.9	3.9 (33)
Z2	F61	M1-M2	p=0.005	2.4	2.3 (46)
		M7-M8	p=0.004	2.4	2.3 (45)
	F3x4	M1-M2	p=0.3	3.0	1.7 (27)
		M7-M8	p=0.04	3.0	1.7 (33)
Z3	F61	M1-M2	p<0.001	2.4	4.5 (30)
		M7-M8	p<0.001	2.4	4.4 (31)
	F3x4	M1-M2	p<0.001	3.1	5.8 (22)
		M7-M8	p<0.001	3.1	5.7 (23)

[a]Only the sequences of Z domain-encoding exons were used in these analyses (see **Figure 2-10** for GenBank accessions).
[b]Two different codon frequency models were used to minimize potentially artificial results.
[c]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.
[d]Tree length provides a measure of nucleotide substitutions per codon along all combined phylogenetic branches.
[e]The percentage of all codons influenced by positive selection is indicated in parentheses.

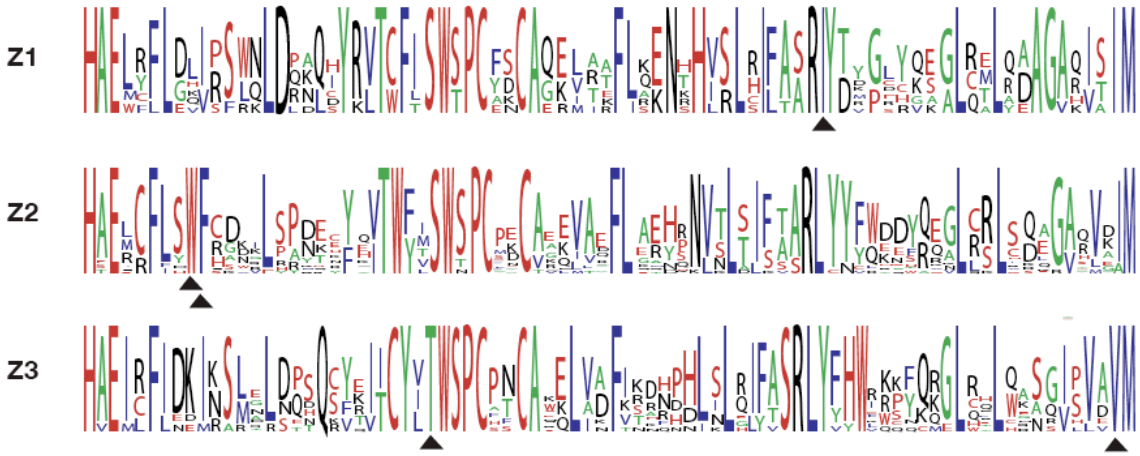


Figure 2-8. APOBEC3 Z domain conservation. Web LOGO profiles depicting amino acid conservation within each mammalian Z domain. The multiple sequence alignments used to generate the phylogenic tree in **Figure 2-1** were used to create consensus profiles for each of the indicated Z domains using Web LOGO (27). Arrowheads below the amino acid profiles indicate residues that define each Z type (see the main text for additional details).

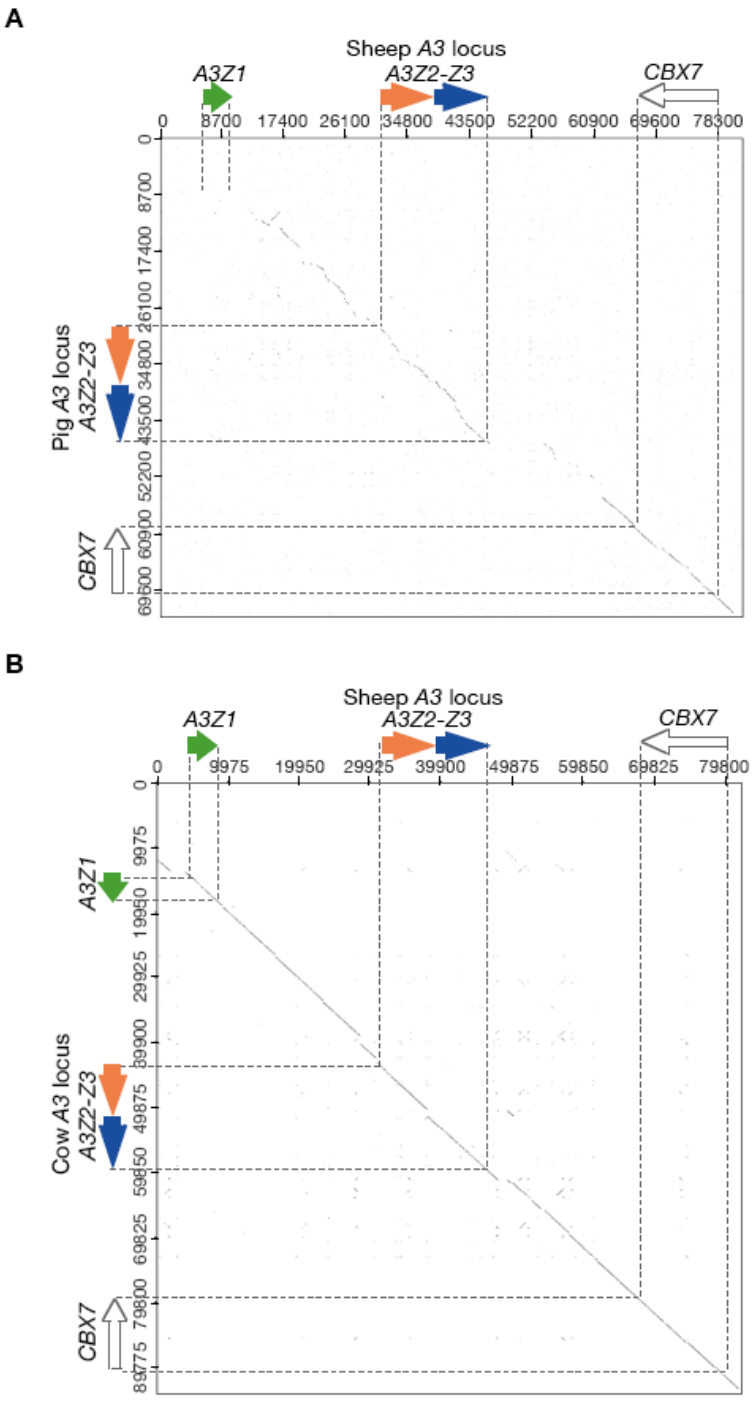


Figure 2-9. *APOBEC3* genomic locus comparisons. Dotplot alignments of **(***A***)** the sheep and pig and **(***B***)** the sheep and cow *A3* genomic sequences.

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

Figure 2-10. Mammalian *A3* and *AID* sequences. A table summarizing all DNA sequences used in this study, including GenBank accession numbers.

A3 gene	ISRE	Distance from ATG
Sheep *A3Z2-Z3, A3FZ2*	TTTACTTTCTCTTTCCCTTT	-311
Cow *A3Z2-Z3, A3Z2*	TTTACTTTCTCTTTCCCTTT	-282
Pig *A3Z2-Z3, 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	

Figure 2-11. *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.

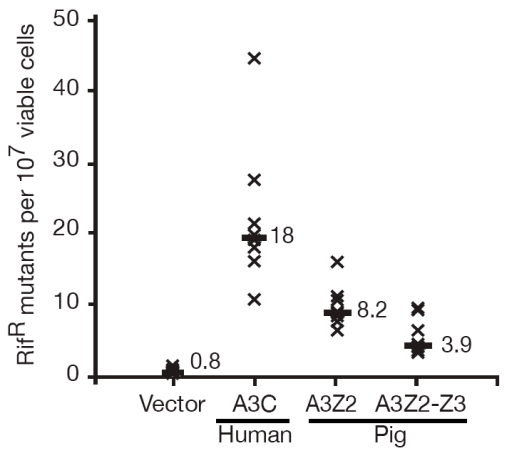


Figure 2-12. *E. coli*-based DNA cytosine deaminase activity data. DNA cytosine deaminase activity of the pig A3Z2-Z3 and A3Z2 proteins in *E. coli*. Conditions and labels are identical to those used in **Figure 2-4**, except 10 independent cultures were grown under IPTG-induced conditions and analyzed.

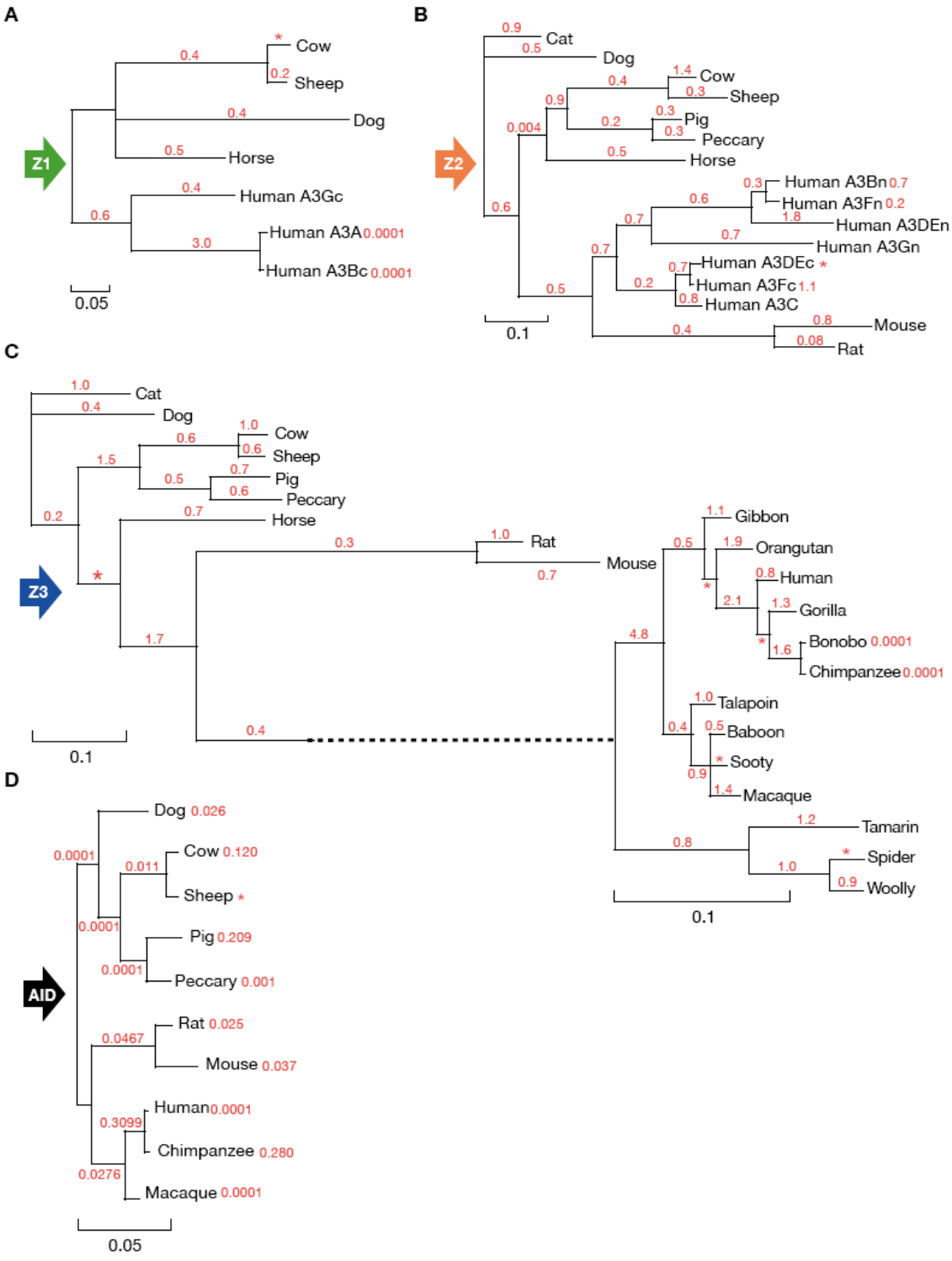


Figure 2-13. Evidence for positive selection in *APOBEC3* gene evolution. Phylogenetic trees showing relative relationships and ω values for the indicated (*A*) Z1 domains, (*B*)

Z2 domains, (*C*) Z3 domains and (*D*) the Z domain of *AID*. The phylogenetic trees were determined using MrBayes, and the ω values were calculated using the PAML free ratio model. ω values are shown in red adjacent to (or where space is non-permitting, to right of) each phylogenetic branch. Asterisks denote branches where the ω 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.

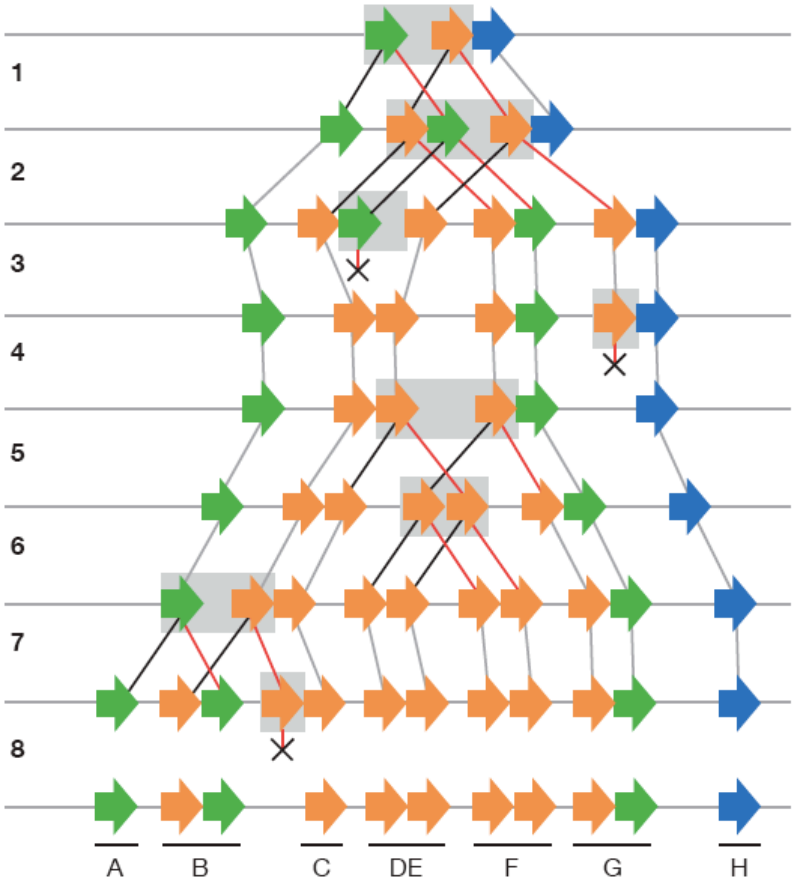


Figure 2-14. Proposed *APOBEC3* gene diversification events during primatification. An alternative representation of the 8-event model for the duplication and deletion history of the human *A3* repertoire. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. The Z domain(s) involved 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 2-7** and **Methods** for details.

	Outer 5' RACE	Inner 5' RACE	Outer 3' RACE	Inner 3' RACE
Cow A3Z1	ACTCTGATGGCACCCAAAAC	GCCTCTCCATCTCGTAGCAC	TTTGGATCAACCGGAGAAAC	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.	TCCGTTCTTGGAATCTGGAC	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.	CCAAGGAGCTGGTTGATTTC	CTGGAGCAATACAGCGAGAG
Pig A3Z2	n.a.	n.a.	ACGTCACCTGGTTCATCTCC	ACTGCTGGAACAACTTCGTG
Pig A3Z3	n.a.	n.a.	Same as pig A3Z2-Z3	Same as pig A3Z2-Z3

n.a. = not applicable

Figure 2-15. Primers used to identify expressed *APOBEC3* transcripts from cow, sheep and pig PBMCs. A table summarizing the oligonucleotide primers used in this study.

CHAPTER 3

Lentiviral Vif Degrades the APOBEC3Z3/APOBEC3H Protein of its Mammalian Host and is Capable of Cross-Species Activity

Reprinted with permission from: R. S. LaRue, J. Lengyel, S.R. Jónsson, V. Andrésdóttir, and R. S. Harris. (2010) J. of Virol.

Authors' contributions

R. S. LaRue and R.S. Harris designed the studies, performed experiments, analyzed data and wrote the manuscript. J. Lengyel performed experiments and analyzed data. S.R. Jónsson and V. Andrésdóttir helped analyze data. All authors contributed to editing the manuscript.

All lentiviruses except EIAV use the small accessory protein Vif to counteract the restriction activity of the relevant APOBEC3 (A3) proteins of their host species. Prior studies have suggested that the Vif-A3 interaction is species specific. Here, using the APOBEC3H (Z3)-type proteins from five distinct mammals, we report that this is generally not the case: some lentiviral Vif proteins are capable of triggering the degradation of both the A3Z3-type protein of their normal host species and those of several other mammals. For instance, SIV_{mac} Vif can mediate the degradation of the human, macaque, and cow A3Z3-type proteins, but not of the sheep or cat A3Z3-type proteins. MVV Vif is similarly promiscuous, degrading not only sheep A3Z3 but also the A3Z3-type proteins of humans, macaques, cows and cats. In contrast to the neutralization capacity of these Vif proteins, HIV, BIV, and FIV Vif appear specific to the A3Z3-type protein of their hosts. We conclude, first, that the Vif-A3Z3 interaction can be promiscuous and, second, despite this tendency, that each lentiviral Vif protein is optimized to degrade the A3Z3 protein of its mammalian host. Our results thereby suggest that the Vif-A3Z3 interaction is relevant to lentivirus biology.

Introduction

Lentiviruses are a unique class of complex retroviruses that encode a variety of accessory proteins in addition to the required Gag, Pol, and Env proteins. The archetypal lentivirus, HIV-1, infects humans, but other members include SIV, BIV, MVV, CAEV, EIAV and FIV, which infect monkeys, cattle, sheep, goats, horses, and cats, respectively. The HIV-1 accessory protein, viral infectivity factor (Vif), has been extensively studied because of its essential function in inhibiting the cellular anti-retroviral human APOBEC3G (A3G) protein (153). HIV-1 Vif binds to human A3G (and other A3 proteins) and serves as an adaptor to link it to an ELOC-based E3 ubiquitin ligase complex (108, 182, 187). A3G is then poly-ubiquitinated and degraded by the cellular proteasome (26, 64, 106, 108, 153, 161, 187).

Due to the potential therapeutic value of disrupting this host-pathogen interaction, a significant amount of work has been invested in defining the important contact residues between A3G and HIV-1 Vif. Primate A3G homologs have been useful tools in this effort as many fail to be neutralized by HIV-1 Vif despite a relatively high degree of sequence similarity. For example, while HIV-1 Vif effectively neutralizes human A3G, it does not neutralize African green monkey A3G or rhesus macaque A3G despite 77% and 75% identity, respectively (104). The differential capacity of the HIV-1 and SIV_{agm} Vif proteins to degrade the A3G proteins of their hosts led to demonstrations that residue 128 is a key determinant: D128 made each A3G protein susceptible to HIV-1 Vif and K128 to SIV_{agm} Vif (147). This apparent on/off switch led to the prevailing model that the Vif-A3 interaction is species specific. However, even early data sets showed at least two hints that the story was more complex. First, the identity of the A3G residue 128 (K or D) does

not diminish the interaction with the Vif proteins of SIV_{mac} or HIV-2 (147, 182). Second, SIV_{mac} Vif was shown to potently counteract the A3G proteins from rhesus macaque (as expected) but also those from human, African green monkey, and chimpanzee (104). Therefore, the implication from these studies is that the full nature of the A3-Vif interaction has yet to be elucidated.

Although A3G has clearly served as the prototype for understanding the A3-Vif interaction, a growing number of studies indicate that other A3s are also capable of restricting lentivirus replication and interacting with Vif. A3G is one of seven human A3 proteins (A3A-H) encoded in tandem on chromosome 22 (26, 66, 177). All but A3A have been implicated in the restriction of HIV-1 replication [reviewed in (3, 51, 156)]. For instance, human A3H has been shown to restrict HIV-1 replication and is susceptible to degradation by HIV-1 Vif (29, 125, 166). A3H is a Z3-type DNA deaminase characterized by a conserved threonine and a valine, in addition to the canonical H-x_1-E-x_{23-28}-C-x_{2-4}-C zinc-coordinating motif (88). The Z3-type deaminase is unique in that only one copy exists in all mammals whose genomes have been sequenced. It is encoded by a five exon gene located at the distal end of each mammal's *A3* locus (adjacent to *CBX7*). Additional observations suggest that the Z3-type deaminases appear to have the capacity to restrict the Vif-deficient lentiviruses of their hosts. For example, African green monkey A3H can restrict the replication of SIV_{agm} and is susceptible to degradation by SIV_{agm} Vif, and the cat A3Z3 can restrict the replication of FIV and is susceptible to degradation by FIV Vif (115, 125, 174).

Here, we take advantage of the fact that all sequenced mammals have a single A3Z3-type protein to test the hypothesis that these proteins are of general relevance to

lentivirus restriction and to clarify the species specific nature of the mammalian A3Z3/lentiviral Vif relationship. First, we ask if human, rhesus macaque, cow, sheep and cat A3Z3-type proteins are all capable of retrovirus restriction. Second, we ask whether they are susceptible to Vif-mediated degradation in a host-specific manner. We show that each lentiviral Vif protein can indeed neutralize the Z3-type A3 protein of its host species. However, we were surprised to find that several of the Vif proteins, particularly SIV_{mac} and MVV Vif, can neutralize a broad number of A3Z3 proteins irrespective of the species of origin and overall degree of similarity. These data indicate that the A3-Vif interaction is more promiscuous than previously appreciated. Such broad functional flexibility may be relevant to understanding past retroviral zoonoses and predicting potential future events. We conclude that the A3Z3-Vif interaction is conserved on a macroscopic level, consistent with an important role in viral replication and particularly in species like artiodactyls and felines with fewer A3 proteins.

Materials and Methods

Sequence alignments. Mammalian A3Z3 protein sequences were aligned and scored using the default matrix BLOSUM62 by protein BLAST (4, 61). A3Z3 sequences compared and tested in this study: human A3H (gb EU861361), rhesus macaque A3H (gb DQ507277), cow A3Z3 (gb EU864536), sheep A3Z3 (gb EU864543) and cat A3Z3 (gb EU011792). Protein sequences for the lentiviral Vif proteins were aligned using T_coffee

version 5.31 (123). All GenBank accession numbers for the lentiviral Vif protein sequences are listed in the Vif expression plasmids section below.

APOBEC3 expression plasmids. The cow A3Z1, A3Z2, A3Z3 and A3Z2-Z3, and sheep A3Z3 genes have been described (68). These cDNAs were subcloned from the pEGFP-N3 expression vector by digesting with KpnI and SalI and inserting them into a similarly digested pcDNA3.1 eukaryotic expression vector with no stop codon. This expression vector added three consecutive HA epitope tags to the CTD of the encoded A3 protein. All of the remaining mammalian A3Z3 genes were also inserted into the pcDNA3.1 expression vector and details are described below. The human A3H (haplotype II) and rhesus A3H cDNAs were kindly provided by Dr. M. Emerman (126) (125). These A3H sequences contained an NTD tag that was removed by PCR amplification with these primers: human forward 5'-NNN NGA GCT CGG TAC CAC CAT GGC TCT GTT AAC AGC CGA AAC-3' and reverse 5'-NNN GTC GAC TCC GGA CTG CTT TAT CCT CTC AAG CC-3' and rhesus forward 5'-NNN NGA GCT CGG TAC CAC CAT GGC TCT GCT AAC AGC-3' and reverse 5'-NNN GTC GAC TCC TCT TGA GTT GCG TAT TGA CGA TG-3'. The untagged human A3H (haplotype II) was amplified with the same forward primer and a reverse primer with a stop codon 5'-NNN NNG TCG ACT CAG GAC TGC TTT ATC CTC TCA AGC CGT C-3'. To ensure the same sequence for the cat A3Z3 gene from (115), RNA was extracted from CRFK cells, reverse transcribed and 3' RACE was performed on the cDNA as described previously (88). The following primers were used to amplify the cat A3Z3 gene without a stop codon: forward 5'-NNN NGA GCT CAG GTA CCA CCA TGA ATC CAC TAC AGG

AA-3’ and reverse 5’- NNN NGT CGA CTT CAA GTT TCA AAT TTC TGA A -3’. The mouse A3Z2-Z3 and human A3G expression plasmids were described (68).

Vif expression plasmids. The lentiviral Vifs chosen for codon optimization (GenScript Corporation) match HIV-1_{IIIB} (gb EU541617), SIV_{mac239} (gb AY588946), BIV_{BIM127} (gb M32690), MVV Icelandic strain 1514 (gb M60610), and FIV_{NSCU} (gb M25381). All of the Vifs were amplified by PCR from the pUC57 vector (GenScript Corporation) and cloned into the pVR1012 vector (Vical Co**.**). The HIV-1_{LAI} Vif in the pVR1012 vector was a generous gift from X.F. Yu, and it is not codon-optimized. All Vifs were amplified by PCR and designed to add a c-myc sequence to the CTD, digested with NotI or SalI and BamHI and cloned into the pVR1012 vector. The following primers were used for the lentiviral Vifs: HIV-1_{IIIB} Vif forward 5’-CAC AAC AAG GTG GGC GCA GCG GCG TAC CTT GCA CTG GCC-3’ and reverse 5’-GGA TCC CTA CAG ATC CTC TTC TGA GAT GAG TTT TTG TTC GTG GCC ATT CAT TGT-3’, HIV-1_{LAI} Vif forward 5’- GTC GAC GCC ACC ATG GAA AAC AGA TGG -3’ and reverse 5’- NNN NGG ATC CCT ACA GAT CCT CTT CTG AGA TGA GTT TTT GTT CGT GTC CAT TCA TTG T -3’, SIV_{mac} Vif forward 5’-AAG TAC CAG GTT CCT GCT GCG GCG TAT CTG GCA CTC AAA-3’ and reverse 5’-GGA TCC CTA CAG ATC CTC TTC TGA GAT GAG TTT TTG TTC GGC CAG GAT ACC CAG-3’, BIV Vif forward 5’-CTC TAC CCC ACG CCA CGC CGC GGC GCG GCT GGC AGC TCT G -3’ and reverse 5’- GGA TCC CTA CAG ATC CTC TTC TGA GAT GAG TTT TTG TTC AGG GTG TCC GCT CAG -3’, MVV Vif forward 5’- AAC ACT AAC CCC AGA GCC GCG GCG AGA CTT GCC CTG CTT -3’ and reverse 5’- GGA TCC CTA CAG ATC CTC

TTC TGA GAT GAG TTT TTG TTC CTC AAA AAT GCT CTC -3', and FIV Vif forward 5'- AAC AGC CCA CCA CAG GCC GCG GCG CGG CTG GCC ATG CTG -3' and reverse 5'-GGA TCC CTA CAG ATC CTC TTC TGA GAT GAG TTT TTG TTC CAG GGA GCC AGA CCA -3'.

Vif mutants were constructed using standard site-directed mutagenesis protocols (Stratagene). The following primers were used on the codon-optimized lentiviral Vifs: HIV-1_{IIIB} Vif (SLQ to AAA) forward 5'-CAC AAC AAG GTG GGC GCA GCG GCG TAC CTT GCA CTG GCC-3' and reverse 5'-GGC CAG TGC AAG GTA CGC CGC TGC GCC CAC CTT GTT GTG-3', SIV_{mac} Vif (SLQ to AAA) forward 5'-AAG TAC CAG GTT CCT GCT GCG GCG TAT CTG GCA CTC AAA-3' and reverse 5'-TTT GAG TGC CAG ATA CGC CGC AGC AGG AAC CTG GTA CTT-3', BIV Vif (SLQ to AAA) forward 5'-CTC TAC CCC ACG CCA CGC CGC GGC GCG GCT GGC AGC TCT G-3' and reverse 5'-CAG AGC TGC CAG CCG CGC CGC GGC GTG GCG TGG GGT AGA G-3', MVV Vif (SLQ to AAA) forward 5'-AAC ACT AAC CCC AGA GCC GCG GCG AGA CTT GCC CTG CTT-3' and reverse 5'-AAG CAG GGC AAG TCT CGC CGC GGC TCT GGG GTT AGT GTT-3', and FIV Vif (TLQ to AAA) forward 5'-AAC AGC CCA CCA CAG GCC GCG GCG CGG CTG GCC ATG CTG-3' and reverse 5'-CAG CAT GGC CAG CCG CGC CGC GGC CTG TGG TGG GCT GTT-3'. The following primers were used for HIV-1_{LAI} Vif (SLQ to AAA): forward 5'-GAC ATA ACA AGG TAG GAG CTG CTG CAT ACT TGG CAC TAG CA-3' and reverse 5'-TGC TAG TGC CAA GTA TGC AGC AGC TCC TAC CTT GTT ATG TC-3'.

HIV-GFP infectivity assays. Human 293T cells were cultured in DMEM supplemented with 10% fetal bovine serum and 0.5% penicillin/streptomycin. At 50% confluency, cells were transfected (TransIt, Mirus) with 1 µg HIV-GFP cocktail [0.44 µg of CS-CG (111), 0.28 µg of pRK5/Pack1(Gag-Pol), 0.14 µg pRK5/Rev, 0.14 µg of pMDG (VSV-G Env)], 100 ng of A3 or empty vector and 20 ng codon-optimized Vif or 400 ng of wildtype HIV-1_{LAI} Vif with corresponding empty vector. The ratio of A3 to Vif constructs was based on whether or not the Vif was codon-optimized. All experiments were performed in triplicate except where noted in the legends.

After 48 hours, virus-containing supernatants were harvested and purified by centrifugation to remove any remaining producer cells. Producer cells were processed for Western analysis and GFP+ flow cytometry (Cell Lab Quanta™ SC-MPL, Beckman Coulter) to measure protein expression levels and transfection efficiencies, respectively. The purified supernatants were placed on fresh 293T target cells. After 72 hours, target cells were harvested and infectivity (GFP) was measured by flow cytometry. Data was analyzed using FlowJo flow cytometry analysis software, version 8.7.1.

Western blotting of cell lysates and viral particles. Producer cell lysates were harvested and pooled from triplicate experiments. Lysates and viral particle pellets were resuspended in 2x Laemmli sample buffer (25 mM Tris pH 6.8, 8% glycerol, 0.8% SDS, 2% 2-mercaptoethanol, 0.02% bromophenol blue), heated at 98°C degrees for 10 min and a fraction of the sample was run on 12% SDS PAGE gel. Protein was transferred to PVDF membrane (Millipore). Membranes with cell lysates were first probed with anti-c-myc (Sigma Aldrich) for detection of Vif expression, then incubated with a horseradish

peroxidase-conjugated secondary antibody and developed using a chemiluminescent HRP antibody detection reagent (HyGLO, Denville Scientific). Blots were then stripped (0.2M glycine, 1.0% SDS, 1.0% Tween-20, pH 2.2) and this protocol was repeated with an anti-HA antibody (Covance) to detect A3 proteins and anti-tubulin (BioRad) to monitor protein loading. To detect human A3H, we used an anti-human monoclonal A3H antibody kindly provided by Dr. M. Emerman (95).

Results

Mammalian A3Z3 proteins are relatively conserved in comparison to lentiviral Vifs. All mammals have only one Z3-type A3 domain (**Figure 3-1A**). We previously described the Z3-type A3 proteins of cattle and sheep and calculated evidence for positive selection in non-human lineages (88). Primate and cat A3Z3 proteins also show evidence for positive selection (115, 126). Accordingly, these mammalian A3Z3 proteins show significant variation, ranging from 48% to 88% identity and 62% to 92% similarity (**Figure 3-1B**). The percent identity and similarity of the mammalian A3Z3 protein sequences correlates with evolutionarily related species, where human and rhesus A3Z3 proteins are 83% identical (90% similar) and cow and sheep A3Z3 proteins are 88% identical (92% similar) (**Figure 3-1A and B**). The cat A3Z3 protein is the most removed ranging from 51% to 52% identity (65% to 66% similarity) with the human and rhesus A3Z3s and 62% identity (74% to 77% similarity) with the cow and sheep A3Z3s. Although A3Z3 proteins (or A3 proteins in general) do not share high sequence identity,

as is the case for essential cellular genes like those that encode the core of the ribosome, there are clearly conserved residues and motifs such as the conserved zinc-binding domain found in all A3 proteins and residues unique to A3Z3 proteins (87).

In contrast, the Vif proteins of lentiviruses are vastly different in primary amino acid sequence. Alignments of known lentiviral Vif protein sequences show less than 30% identity, even between evolutionarily related viruses (124, 151). However, despite low sequence identity between different Vifs, the proteins share a highly conserved S/TLQY/RLA motif (**Figure 3-1C**) (124, 151). This virus-specific motif in HIV-1, HIV-2, different primate SIVs, BIV and MVV Vifs is essential for binding ELOC, which is part of a ubiquitin ligation complex (ELOB/C, CUL2 or CUL5, and RBX) capable of degrading A3s (100). The primate lentiviral Vifs (HIV and SIV) are unique from the other mammalian Vifs in that they share a zinc-binding domain $HX_5CX_{17-18}CX_{3-5}H$ (HCCH) that binds to CUL5 (100). BIV and MVV Vif lack this domain and therefore do not co-immunoprecipitate with the CUL5 protein when expressed in human cells (100). MVV Vif, on the other hand, co-immunoprecipitates the CUL2 protein indicating that non-human Vifs may utilize a different Cullin (100). Due to broad conservation of the Vif S/TLQY/RLA motif, we constructed triple amino acid substitution derivatives (S/TLQ to AAA) of each lentiviral Vif to use as controls. This mutant is unable to bind ELOC and it is a common negative control for many HIV-1 and SIV Vif studies [*e.g.*, (77, 100, 154, 187)].

Representative mammalian A3Z3 proteins can restrict HIV. Every completely annotated mammalian *A3* locus to date has only one copy of a Z3-type *A3* gene (*A3H-*

like). We hypothesized that all A3Z3-type proteins are neutralized by their host-specific lentiviral Vif. A3Z3 proteins from human, rhesus macaque and cat were selected because these mammals are infected with a Vif-containing lentivirus and are capable of restricting HIV or its host lentivirus in the absence of a functional Vif protein (29, 31, 53, 95, 115, 125, 174). The A3Z3 proteins of cow and sheep were also chosen for the representative panel because these artiodactyls are also infected with Vif-containing lentiviruses. Previous studies in our lab have shown that the cow and sheep A3Z2-Z3 can restrict HIV (68). The cow and sheep A3 proteins are also naturally expressed as single domain variants (88), making them more ideal for comparisons with the Z3-type A3s of other mammals.

First, we performed a series of single-cycle HIV-GFP infectivity assays to test whether the single domain cow A3 proteins (A3Z1, A3Z2 and A3Z3) could restrict an HIV-based virus construct. The HIV-GFP viral cocktail was transfected into 293T cells alongside an empty pcDNA3.1 vector or one of the cow A3 proteins in the presence of empty pVR012 vector, BIV Vif-myc, or the BIV $Vif_{SLQ\text{-}AAA}$-myc. A portion of the producer cell virus-containing supernatant was used to infect fresh 293T cells to measure infectivity by GFP+ flow cytometry. We were surprised that HIV-GFP was restricted equally well by the cow single domain A3Z3 protein and the double domain A3Z2-Z3 protein (**Figure 3-2A**). The cow A3Z1 protein showed no anti-HIV activity, despite being a very active DNA deaminase (88). As controls used previously by many labs, human A3G and mouse A3 both restrict the infectivity of HIV-GFP [*e.g.*, (68)].

In parallel experiments, we examined the Vif susceptibility of the cow A3 proteins and found that BIV Vif restores the infectivity of viruses produced in the

presence of cow A3Z3 and A3Z2-Z3, but not those produced with human A3G or mouse A3 (**Figure 3-2A**). These data suggested that BIV Vif might only interact with the cow Z3-type deaminase domain. However, immunoblots of the proteins expressed in the producer cell extracts showed that the Z2-type domain was also clearly susceptible to Vif-mediated degradation despite the fact that this domain is not anti-viral (**Figure 3-2B**).

As anticipated by the infectivity data, BIV Vif also degraded the cow A3Z3 and A3Z2-Z3 proteins. None of the cow A3s are degraded in the presence of mutant BIV $Vif_{SLQ\text{-}AAA}$-myc (**Figure 3-2**). Overall, these experiments demonstrated that BIV Vif triggers the degradation of both Z2- and Z3-type A3s and that Vif function requires the conserved SLQ motif (and therefore likely also the interaction with ELOC). In addition, the results show that BIV Vif either does not interact with or does not trigger the degradation of the Z1-type A3.

BIV Vif function appears species specific in contrast to that of the related MVV Vif protein. We next asked whether BIV Vif could neutralize other mammalian A3Z3 proteins besides cow A3Z3. HIV-GFP was produced in the presence of a diverse set of mammalian A3Z3 proteins and, as above, the producer cells were analyzed by immunoblotting and the resulting viral supernatants by infecting susceptible target cells. The A3Z3 proteins showed a range of HIV-GFP restriction phenotypes with rhesus macaque A3H appearing to have the most potency and the sheep A3Z3 the least (**Figure 3-3A**). Some of these differences may be due in part to slight variations in protein expression. However, consistent with the species specific hypothesis, BIV Vif only rescued the infectivity of viruses produced in the presence of cow A3Z3. This counter-

restriction effect correlated with greatly reduced cellular cow A3Z3 protein levels. In contrast, even the sheep A3Z3, which is 92% similar to the cow A3Z3, showed no sign of being degraded.

To ask whether these observations would be reciprocal and possibly generalizable, we turned to the most closely related lentivirus-host interaction between MVV Vif and the sheep A3s. The predicted evolutionary distance between the sheep and cow is similar to that between the human and rhesus macaque, approximately 28 million years (**Figure 3-1A**) (9). BIV and MVV have similar genome organizations (124) and it is reasonable to presume that these viruses arose from a common ancestor (49, 73). However, in comparison to the species specific activity of BIV Vif, MVV Vif counteracted the restriction activity of nearly all of the A3Z3 proteins (**Figure 3-3B**). MVV Vif neutralized the sheep and cow A3Z3 proteins with near 100% efficiency, and it was also very effective at degrading the human, rhesus macaque and cat A3Z3 proteins. These effects were equally apparent at the protein level in immunoblots of producer cell extracts. Interestingly, despite MVV Vif having no effect on the restriction capability of human A3G and the mouse A3, protein degradation was also detected for these intended non-single domain A3Z3 controls (human A3G=Z2Z1 and mouse=Z2Z3). These phenotypes are not observed with the MVV $Vif_{SLQ\text{-}AAA}$-myc construct but this maybe due to the apparent instability caused by the SLQ mutations (**Figure 3-3B**). Taken together, these data demonstrate that MVV Vif lacks both species specificity and Z-type specificity, based on its capability of triggering the degradation of a broad array of mammalian A3 proteins that contain different combinations of the Z-type domains.

MVV Vif-mediated degradation of the A3s is not simply an artifact of degrading all cellular proteins because tubulin levels were unaffected and toxicity was not observed (**Figure 3-3B** and data not shown). Such promiscuity was unanticipated, so we next asked whether MVV Vif has a preference for degrading sheep A3Z3. We analyzed the infectivity of HIV-GFP produced in the presence of sheep A3Z3, human A3H, mouse A3, or human A3G and range of MVV Vif levels (0, 5, 20, and 50 ng plasmid per transfection). Even the lowest MVV Vif levels were sufficient to rescue HIV-GFP infectivity by counteracting sheep A3Z3 or human A3H (data not shown). However, immunoblot analysis of the producer cell lysates showed that only the sheep A3Z3 was efficiently degraded at the lowest concentrations, suggesting that MVV Vif has some preference for sheep A3Z3.

HIV-1 Vif neutralization of human A3H and other mammalian A3Z3 proteins. Prior reports have indicated that HIV-1 Vif fails to neutralize rhesus macaque A3H or cat A3Z3 (116, 174). Recent studies with HIV-1 Vif and human A3H (haplotype II) have yielded conflicting results, with some showing neutralization and others not (29, 31, 53, 95, 125). To further investigate the potential interaction between HIV-1 Vif and human A3H and, if apparent, ask whether it is species specific, we did a series of HIV-GFP infection experiments as described above with HIV-1 Vif from the LAI strain and our panel of mammalian A3Z3s. As observed in a prior study with HIV-1_{LAI} Vif (95), HIV-GFP infectivity was partly restored and the level of cellular human A3H slightly diminished (**Figure 3-4A**). HIV-GFP infectivity in the presence of the cow, sheep and mouse A3s was also improved significantly by expressing HIV-1_{LAI} Vif, but this effect

was not accompanied by a significant change in cellular protein levels. It is unclear whether all of these effects were dependent on the conserved SLQ motif because the mutated protein did not express well in comparison to the wildtype protein. Overall, however, the phenotypes attributable to HIV-1_{LAI} Vif were quite modest in comparison to results with other lentiviral Vif proteins or data with HIV-1 Vif and human A3G [**Figure 3-2**, **Figure 3-3**, below, and prior studies, *e.g.*, (25, 106, 108, 154, 161, 187, 188)].

We therefore wondered whether the C-terminal HA epitope tag might be protecting human A3H from Vif-mediated degradation. To address this point, we compared the sensitivity of human A3H-HA and untagged A3H to neutralization by HIV-1_{LAI} Vif (**Figure 3-4B**). In both instances, the infectivity of HIV-GFP recovered significantly and this effect corresponded to the near disappearance of the A3H-specific band from the cell lysate immunoblot. A similarly large amount of A3H degradation was apparent in analogous experiments with HIV Vif from strain IIIB (data not shown). The only significant difference between the immunoblot experiments shown in **Figure 3-4A** and those in **Figure 3-4B** is the fact that the former required the use of an anti-HA monoclonal antibody to compare all of the A3s and the latter used a A3H-specific monoclonal antibody to detect the human protein [kindly provided by M. Emerman (95)]. To further investigate this suspected difference in antibody sensitivity, the blot in **Figure 3-4B** was stripped and probed with the anti-HA antibody and a faint band was detected for the HA-tagged A3H expressed with HIV Vif (data not shown). So the difference in A3 degradation is most likely linked to antibody sensitivity (with the anti-HA monoclonal being more sensitive). Nevertheless, the sum of our results combined to indicate that HIV-1 Vif has a notable preference for human A3H over other Z3-type A3s.

The modest neutralization of two of the other A3s may be related to the degradation-independent inhibition noted previously for HIV-1 Vif (72, 128, 144).

SIV_{mac} Vif and FIV Vif differ in their ability to neutralize other species A3Z3 proteins. Many previous studies have investigated the primate A3 neutralization capabilities of SIV_{mac} Vif (104). SIV_{mac} Vif is capable of degrading rhesus macaque A3H, African green monkey A3H, and human A3H (174). To extend these studies, we tested SIV_{mac} Vif against our full panel of mammalian Z3-type A3 proteins (**Figure 3-5A**). In addition to counteracting the restrictive capabilities of the rhesus and human A3H proteins, SIV_{mac} Vif neutralized cow A3Z3. Again, a corresponding drop in cellular A3 levels was observed. However, SIV_{mac} Vif failed to neutralize the sheep A3Z3 protein (surprising since it is 92% similar to the cow protein) and the cat and the mouse proteins. Thus, SIV_{mac} Vif is not quite as promiscuous as the MVV Vif, but it is clearly more so than the BIV or the HIV-1 Vif proteins.

These experiments also revealed another curious phenotype, as the SLQ motif of SIV_{mac} Vif appeared partly dispensable for its counter-restriction effect against the human and rhesus A3H proteins (**Figure 3-5A**). Together with the aforementioned studies, it is possible that Vif is capable of eliciting two degradation-independent effects – an ELOC-independent mechanism as observed here with SIV_{mac} Vif and an ELOC-dependent mechanism as observed in **Figure 3-4A** with HIV-1 Vif. Further studies may be warranted to better understand this potentially degradation-independent mechanism of A3 neutralization.

Finally, to complete the survey we tested the sensitivity of the mammalian A3Z3 panel to FIV Vif (**Figure 3-5B**). These results were remarkably parallel to the BIV Vif data, as only the cat A3Z3 protein was neutralized and degraded by FIV Vif. A similar observation was made previously (115). However, our studies uniquely show that FIV Vif-mediated neutralization of the cat A3Z3 protein is dependent on the TLQ motif, indicating a requirement for the ELOC interaction.

Discussion

In contrast to Z1-type A3s, which have not been reported to be degraded by Vif, and Z2-type A3s, which are sometimes degraded by Vif, we have shown here that Z3-type mammalian A3s are invariably degraded by the Vif protein of each species' lentivirus. Vif-mediated degradation of the Z3-type A3s is also accompanied by improved viral infectivity. This strong correlation provides good evidence that the Vif-A3Z3 interaction is relevant to lentiviral replication (*i.e.*, the lentivirus would not have evolved and retained Vif if these mammalian A3Z3s or their ancestral equivalents were not a significant threat). Moreover, our data indicate that the S/TLQ motif of BIV, SIV_{mac} and FIV Vif is required for degradation of the Z3-type A3s. These data demonstrate that interaction with ELOC is conserved and that the likely mechanism is ubiquitin-mediated degradation. Although the SLQ motif in HIV Vif has been shown to be important for the interaction with ELOC (188), further studies are needed to verify whether this motif is important for HIV_{LAI} Vif and MVV Vif because mutating these residues greatly reduced expression of the mutant forms. The conservation of the Vif-A3Z3 interaction also

suggests the existence of at least one common underlying structural determinant, because the lentiviral Vifs are much more divergent at the amino acid level than the A3Z3s of their mammalian host (**Figure 3-1**).

A second major conclusion to stem from our studies is that some lentiviral Vif proteins lack species specificity (summarized in **Table 3-1**). The best example is MVV Vif, which triggers the near complete degradation of the sheep, cow and cat A3Z3 proteins and partial degradation of the human and rhesus macaque A3H proteins. MVV Vif also restores or improves HIV infectivity in the presence of each mammalian A3Z3 protein. In contrast, the closest relative to this protein, BIV Vif, only appears to neutralize the cow A3Z3 protein. A second example is SIV_{mac} Vif, which triggers the degradation of the rhesus macaque, human and cow A3Z3 proteins and provides a corresponding improvement in virus infectivity. Again, in contrast, the closest relative to this protein, HIV-1 Vif, appears fairly specific to the human A3H protein. It is probable that had we expanded the survey to include an even broader range of mammalian A3s, that each lentiviral Vif would be able to trigger the degradation of multiple A3s (*i.e.*, the Vif-A3Z3 interaction appears far from species specific). It is further probable that Vifs from different lentiviral strains within a species will possess a range of A3 neutralization capabilities as has been indicated by the different strains of HIV-1 (105).

Another apparent trend in our data set is that Vif may have the capacity to antagonize the restriction activity of a given A3Z3 either without triggering its degradation or by utilizing a different degradation pathway. First, we observed an increase in HIV-GFP infectivity when HIV-1_{LAI} Vif is co-expressed with cow A3Z3, sheep A3Z3 or mouse A3Z2-Z3, even in the absence of detectable protein degradation. It

may be possible that HIV-1 Vif is interacting in a way that hinders the A3 protein's ability to restrict but is not efficiently recruiting or correctly positioning the cellular degradation machinery complex (72, 128, 144). Second, we observed that the SIV_{mac} $Vif_{SLQ-AAA}$ mutant is capable of triggering partial degradation of the human and rhesus A3H proteins. It is therefore possible that Vif may bind to A3Z3 proteins and recruit an alternative cellular complex to induce degradation. Other possibilities are that degradation may depend in part on species specific interactions or that the Vif variants analyzed here may not be fully representative of the many other variants that exist naturally.

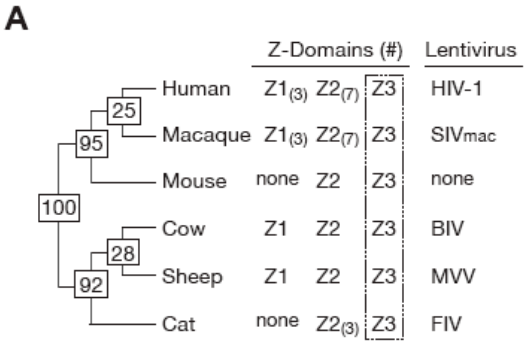
The A3Z3-type deaminase domain can be expressed on its own and show antiviral activity. It can also be expressed as part of a double-domain deaminase (Z2-Z3) in many mammals including cattle, sheep, pigs, cats and mice. The cow, sheep and cat A3Z2-Z3 proteins all restrict HIV and/or FIV and are neutralized by its host-specific Vif (68). We show in this study that the cow A3Z2 protein is degraded in the presence of BIV Vif, even though it does not restrict HIV (**Figure 3-2**). It is therefore possible that Vif is able to simultaneously interact with both types of Z-domains (the A3Z2-Z3 protein), resulting in a synergistic neutralization effect. Additional studies will be needed to systematically address this possibility and test alternatives such as additive neutralization or competitive binding.

Finally, our data clearly demonstrate that the Vif-A3Z3 interaction is not 'species specific' as commonly inferred, but more likely to be host-optimized and considerably promiscuous. The idea of a species specific A3-Vif interaction arose following seminal work that mapped a key residue in A3G, D128, which appeared to dictate the

susceptibility of this protein to Vif-mediated degradation (104). Here, our broad survey suggests that lentiviral Vifs are optimized to interact with their host species' A3Z3 proteins because in all instances the lentiviral Vifs can effectively degrade and neutralize the A3Z3 of their host. However, we also observe that this interaction can be highly versatile and, in multiple instances, Vif is able to recognize and degrade A3 proteins of other species, some from which it is far removed (at least in evolutionary terms-Figure 1A). Therefore, while Vif invariably functions as an A3 antagonist in its host species, its ability to neutralize A3s of other species provides a mechanism to overcome this key barrier to zoonotic transmission. Additional comparative studies are likely to help map critical interacting regions and inform efforts to develop novel anti-viral drugs that leverage the Vif-A3 axis.

Acknowledgements

We thank J. Albin, J. Hultquist and V. Pathak for comments on the manuscript, M. Emerman for key reagents, X-F. Yu for sharing unpublished data, and J. Weaver for providing technical assistance. R. LaRue is a student in the Comparative Molecular Bioscience program and is supported in part by a studentship from the MN Agricultural Experiment Station (MAES) and College of Veterinary Medicine (CVM). This work was funded by a grant from the National Institutes of Health AI064046. Lastly, we dedicate this paper to Dr. David Derse, who enthusiastically supported our non-human A3 studies over the years.



B

A3Z3 Percent Identity

A3Z3 Percent Similarity	Hs	Rh	Bt	Oa	Fc
Hs		83	48	49	51
Rh	90		51	51	52
Bt	66	67		88	62
Oa	66	62	92		62
Fc	65	66	74	77	

C

Lentiviral Vif Comparison

Vif	Strain	Total # of aa	Conserved Elongin BC Motif
HIV-1	LAI	192	144—SLQYLA—149
SIV	mac239	214	147—SLQYLA—152
BIV	p127	198	150—SLQRLA—155
MVV	Iceland	230	173—SLQRLA—178
FIV	NCSU	251	200—TLQRLA—205
Consensus Motif			SLQYLA T R

Figure 3-1. Comparison of A3Z3 and Vif sequences. (A) The relatedness of the mammals relevant to this study. The numbers at the branch nodes are approximate divergence times in millions of years. Adapted from LaRue et al. (87) (B) Percent sequence identity and similarity between all the mammalian A3Z3 proteins. Abbreviations for different mammals: hs, human; rh, rhesus macaque; bt, cow; oa, sheep; and fc, cat. (C) A alignment of the ELOC binding motif of the lentiviral Vifs used in this study.

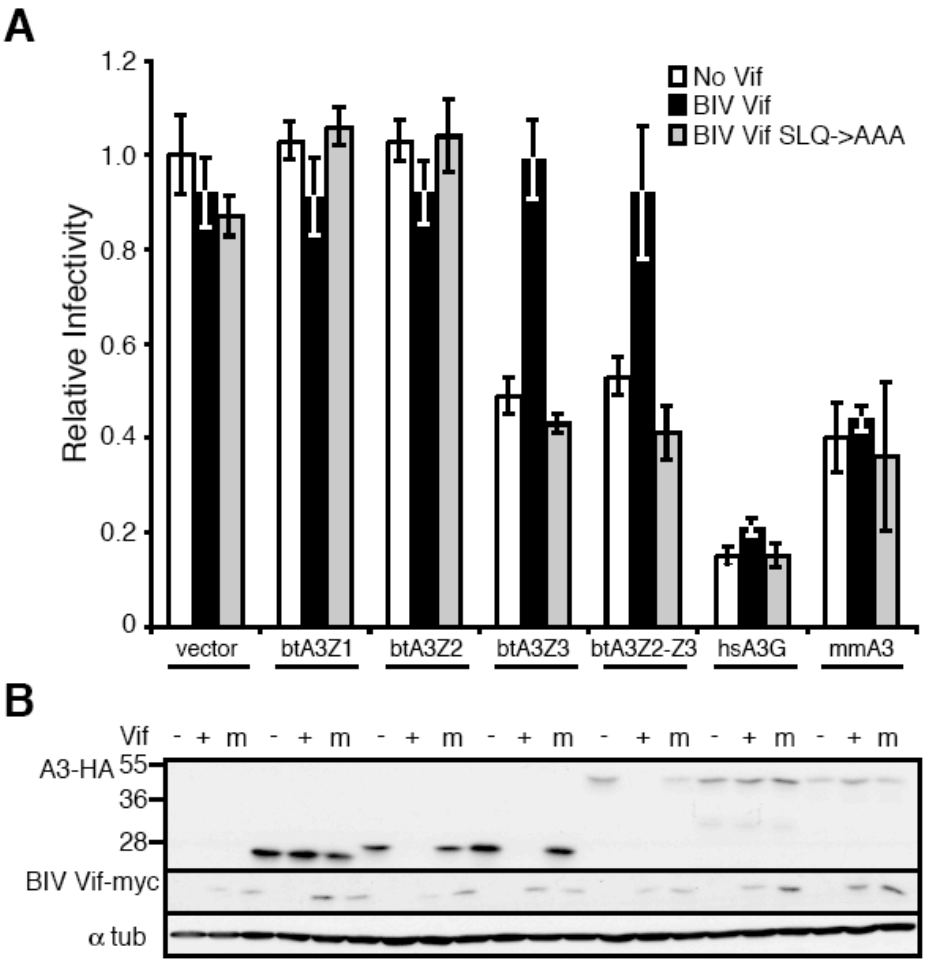


Figure 3-2. Anti-retroviral properties of cow A3 proteins and their sensitivity to BIV Vif. (A) The relative infectivity of HIV-GFP produced in the presence of the indicated A3s and a vector control (open bars), BIV Vif (black bars), or BIV $Vif_{SLQ\text{-}AAA}$ (grey bars). The double vector control (empty pcDNA3.1 and pVR1012 vectors) was normalized to one and the error bars report the SEM from 3 independent experiments. (B) Immunoblots showing the producer cell levels of A3 proteins (anti-HA), Vif (anti-myc), and tubulin.

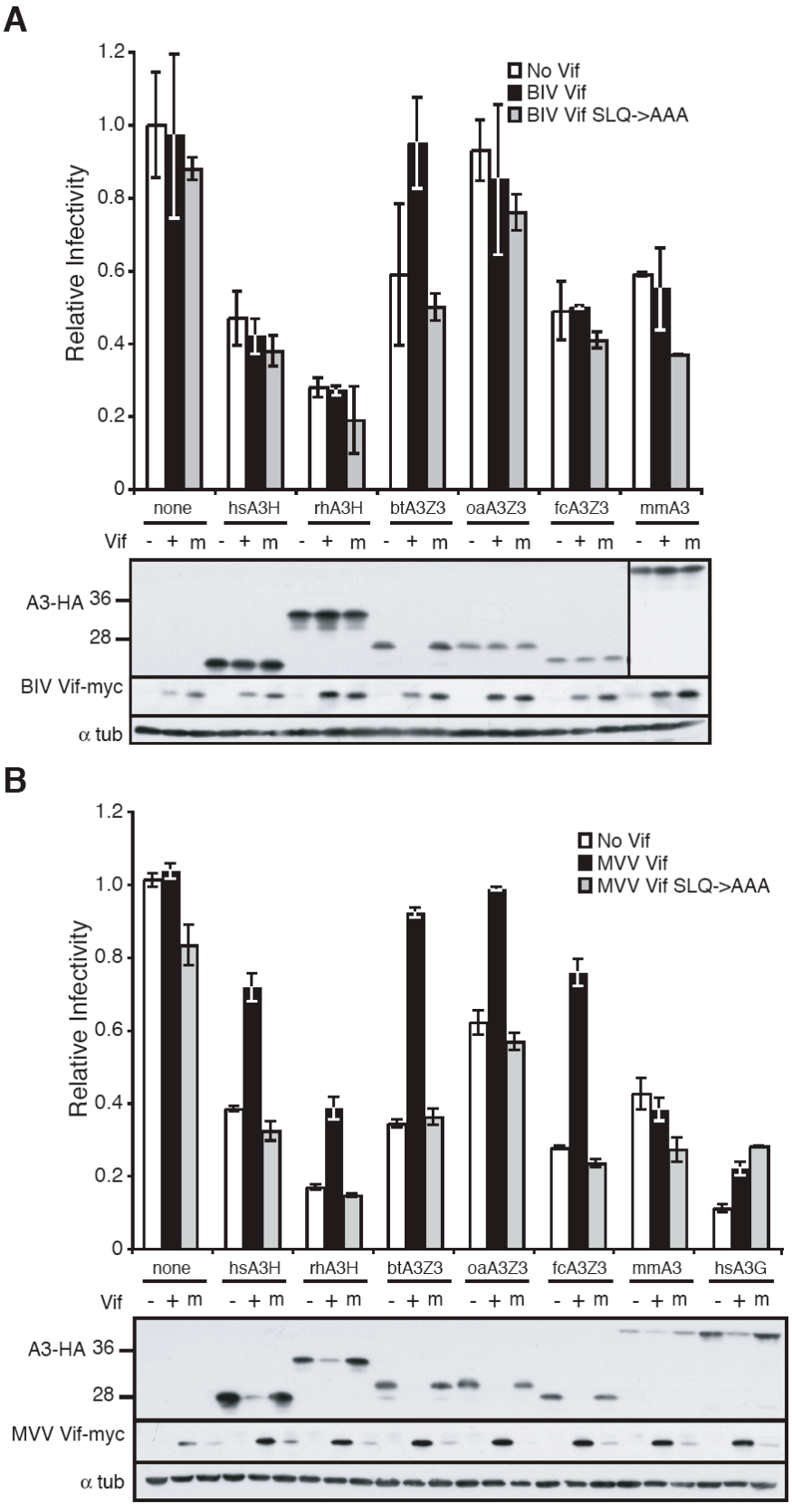

A
Relative Infectivity
1.2
1.0
0.8
0.6
0.4
0.2
0
No Vif
BIV Vif
BIV Vif SLQ->AAA
none
hsA3H
rhA3H
btA3Z3
oaA3Z3
fcA3Z3
mmA3
Vif - + m - + m - + m - + m - + m - + m - + m
A3-HA 36
28
BIV Vif-myc
α tub
B
Relative Infectivity
1.2
1.0
0.8
0.6
0.4
0.2
0
No Vif
MVV Vif
MVV Vif SLQ->AAA
none
hsA3H
rhA3H
btA3Z3
oaA3Z3
fcA3Z3
mmA3
hsA3G
Vif - + m - + m - + m - + m - + m - + m - + m - + m
A3-HA 36
28
MVV Vif-myc
α tub

Figure 3-3. Sensitivity of mammalian A3Z3 proteins to BIV and MVV Vif. (A) The relative infectivity of HIV-GFP produced in the presence of the indicated A3s and a vector control (open bars), BIV Vif (black bars), or BIV $Vif_{SLQ-AAA}$ (grey bars). The double vector control was normalized to one and the error bars report the SEM from 3 independent experiments. Producer cell immunoblots are shown below for A3 proteins (anti-HA), Vif (anti-myc), and tubulin. The mouse A3 signals were faint on the original full immunoblot, so an image of a longer exposure was inserted in the identical position for presentation. (B) The relative infectivity of HIV-GFP produced in the presence of the indicated A3s and a vector control (open bars), MVV Vif (black bars), or MVV $Vif_{SLQ-AAA}$ (grey bars). Organized and labeled as above.

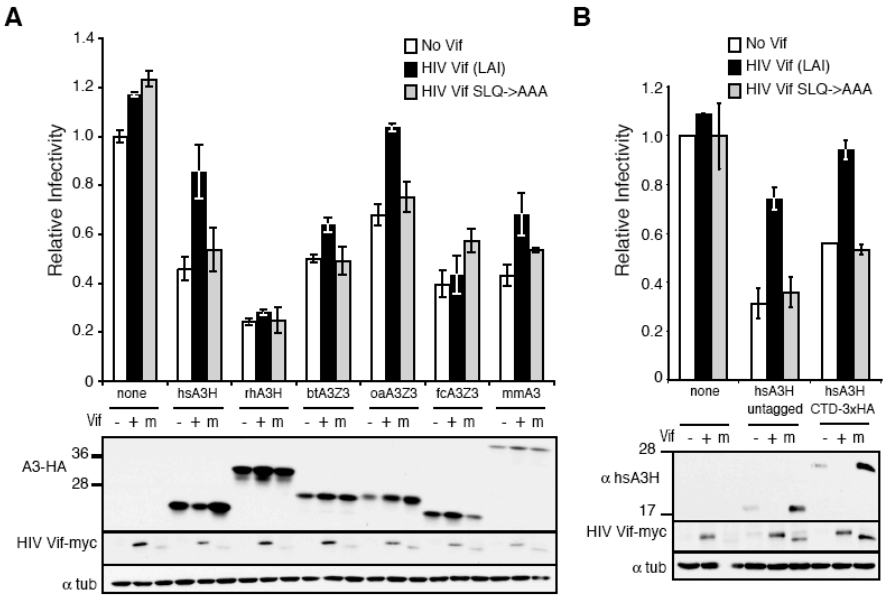


Figure 3-4. Sensitivity of mammalian A3Z3 proteins to HIV-1_{LAI} Vif. (A) The relative infectivity of HIV-GFP produced in the presence of the indicated A3s and a vector control (open bars), HIV-1_{LAI} Vif (black bars), or HIV-1_{LAI} $Vif_{SLQ-AAA}$ (grey bars). The double vector control was normalized to one and the error bars report the SEM from 3 independent experiments. Producer cell immunoblots are shown below for A3 proteins (anti-HA), Vif (anti-myc), and tubulin. (B) The relative infectivity of HIV-GFP produced in the presence of untagged human A3H (haplotype II) or A3H-HA and a vector control (open bars), HIV-1_{LAI} Vif (black bars), or HIV-1_{LAI} $Vif_{SLQ-AAA}$ (grey bars). Producer cell immunoblots are shown below for A3H (anti-A3H), Vif (anti-myc), and tubulin.

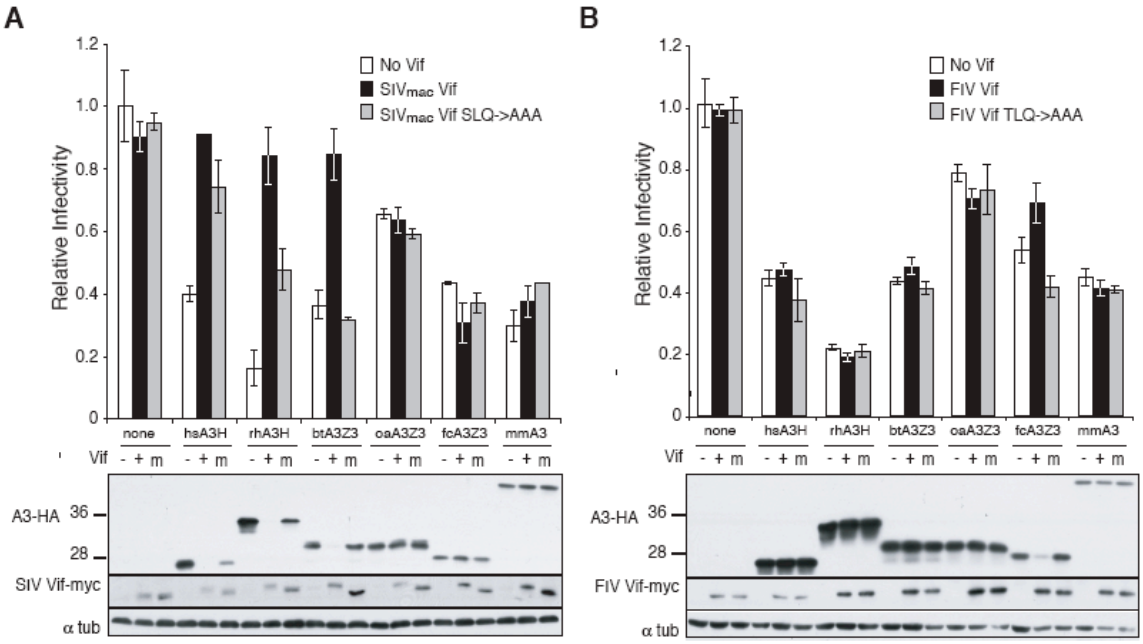


Figure 3-5. Sensitivity of mammalian A3Z3 proteins to SIV_{mac} and FIV Vif. (A) The relative infectivity of HIV-GFP produced in the presence of the indicated A3s and a vector control (open bars), SIV_{mac}Vif (black bars), or SIV_{mac} $Vif_{SLQ-AAA}$ (grey bars). The double vector control was normalized to one and the error bars report the SEM from 3 independent experiments. Producer cell immunoblots are shown below for A3 proteins (anti-HA), Vif (anti-myc), and tubulin. (B) The relative infectivity of HIV-GFP produced in the presence of the indicated A3s and a vector control (open bars), FIV Vif (black bars), or FIV $Vif_{TLQ-AAA}$ (grey bars). Labeled as above. We note that the transfer of protein on this particular blot is slightly less efficient toward the right, but this does not alter our overall conclusions (table 2-1).

Table 3-1. A qualitative summary of Vif-mediated A3 degradation and rescue of HIV-GFP infectivity from A3 restriction.

Degradation[a]	btA3Z3	oaA3Z3	hsA3H	rhA3H	fcA3Z3	mmA3	hsA3G
BIV	++	-	-	-	-	-	-
MVV	++	++	++	++	++	+	++
HIV-1	-	-	+	-	-	-	++[d]
SIVmac	++	-	++	++	-	-	ND[c]
FIV	+	+	-	-	++	-	ND

Rescue of infection[b]	btA3Z3	oaA3Z3	hsA3H	rhA3H	fcA3Z3	mmA3	hsA3G
BIV	**	-	-	-	-	-	-
MVV	**	**	*	*	*	-	-
HIV-1	*	**	**	-	-	*	**[d]
SIVmac	**	-	**	**	-	-	ND
FIV	-	-	-	-	*	-	ND

[a]A3 degradation: ++, 50-100%; +, 5-49%; -, 0-4.9%
[b]HIV infectivity: **, complete rescue; *, partial rescue; -, no rescue
[c]ND, not done
[d]Data not shown

Chapter 4

Conclusions and Discussion

Our analyses of representative artiodactyl *A3* loci have yielded striking differences between the primate and artiodactyl repertoires. As shown in Chapter 2, some artiodactyl *A3* loci encode approximately 2-fold fewer A3 proteins than the characterized primate *A3* loci. Nevertheless, the artiodactyl proteins demonstrate intrinsic DNA cytosine deaminase activity and subcellular localization similar to analogous human A3 proteins with conserved Z-type domains. The findings in this particular study establish a foundation for understanding the evolutionary history of mammalian *A3* genes.

To begin to characterize the artiodactyl repertoire functionally, the cow A3 proteins were individually tested for their ability to restrict a representative lentivirus, HIV. As shown in Chapter 3, the cow A3Z2-Z3 and A3Z3 proteins restrict Vif-deficient HIV whereas the A3Z1 and A3Z2 proteins do not. In the presence of Vif from the cow specific lentivirus BIV, the A3Z2-Z3, A3Z3, and the non-restricting A3Z2 are all neutralized. The rest of Chapter 3 is dedicated to utilizing the conserved anti-retroviral, single domain, Z3-type A3 protein expressed in humans, rhesus macaques, cattle, sheep and cats, to investigate if and at what level the mammalian A3/lentiviral Vif interaction is conserved. All of these mammalian A3Z3 proteins are neutralized in the presence of their species specific Vif. Moreover, Vifs from SIVmac and MVV are able to neutralize other species A3Z3 proteins. The data collected from these comparative experiments are supportive of a conserved interplay between mammalian A3Z3 proteins and lentiviral Vif proteins.

This last chapter will discuss unanswered questions from these studies and the status of my working hypothesis.

The mammalian A3 locus is dynamic

From an evolutionary perspective, the relatively rapid and drastic expansion of the *A3* locus in placental mammals is quite intriguing. All mammals (placentals, marsupials, and monotremes) contain the APOBEC family members *AID*, *APOBEC1*, *APOBEC2*, and *APOBEC4* (**Figure 4-1**). *APOBEC3* genes, however, are unique to placental mammals and are syntenic, located between the two polycomb-like chromobox genes, *CBX6* and *CBX7* (**Figure 4-1**). At some point during the early radiation of placental mammals, an ancestral *A3* gene, hypothesized to be *AID*-like (26, 57), transposed into this location. Subsequent duplication events produced more *A3* genes that evolved into Z1, Z2 and Z3-types (**Figure 4-2A)**. After the split that created the four major clades of placental mammals (**Figure 4-1**), the *A3* loci in the lineages containing primates and horses underwent large expansions whereas the *A3* loci in the lineages containing artiodactyls, felines and rodents maintained a smaller set of ancestral genes (**Figure 4-2B**). The major unanswered question, then, is what selected for such large expansions of *A3* genes in certain mammalian lineages?

Perhaps these large expansions were maintained due to functional differences between the Z-domains (Z1, Z2, Z3). Each type of Z-domain is expressed as single domain protein in all mammals with the exception of those that lack the *Z1* gene (cats and pigs) and the mouse, which only expresses the double-domain A3Z2-Z3 [LaRue, unpublished data and (36)]. The Z2-Z3 organization of the double domain protein is

present in cattle, sheep, pigs and cats (115, 116) as well as rodents (26, 57). This Z-domain organization is not present in primates or horses, which encode double-domain A3 proteins with either a Z2-Z1 organization (such as human A3B and A3G) or a Z2-Z2 organization (such as human A3DE, A3F or horse A3Z2a-Z2b, A3Z2c-Z2d) (12, 195)(**Figure 4-2B**). All of these double-domain A3 proteins have been shown to be capable of restricting a range of different retroelements, but it is unclear how the organization of the Z-domains influences the target and degree of restriction.

It is important to note that the function of A3 proteins from non-primate mammals has largely been characterized with experimental assays that test for intrinsic DNA mutation in *E. coli* [Chapter 2(68, 88)] or restriction of reporter retrovirus produced using the human HEK-293 cell line [Chapter 3(12, 32, 68, 78, 89, 98, 104, 115, 116, 143, 195)]. The functional efficacy of these proteins in the context of their host-derived cell lines and against host-specific retroelements may be quite different. For instance, human A3G can form high molecular weight (HMM) ribonucleoprotein (RNP) complexes that contain at least 95 different proteins [reviewed in (22)]. In this HMM complex, A3G is not able to restrict Vif-deficient HIV, though the exact nature of this regulatory mechanism is unknown. It is unclear if it the pertinent regulatory components are conserved and functional in heterologous systems. Until these post-translational mechanisms controlling function are better understood, expression of these A3 proteins outside of their normal cellular context may alter their function in unpredictable ways and therefore the conclusions drawn from such systems must be evaluated cautiously. Nevertheless our functional data from this system is supported by evidence that, like the primate lineages, the artiodactyl A3 lineages are also under positive selection. This would

be predicted if the primary role of these A3 proteins is, as in primates, to protect the genome from exogenous and endogenous pathogens.

Another factor to consider towards determining the function of these artiodactyl A3s will be the expression profile in different tissues of the host. With the exception of human A3 proteins (58, 79, 97, 135), one study involving horse A3 proteins (12), and one study involving the mouse A3Z2-Z3 protein (36), not much is known about the A3 expression in tissues of non-human primates, artiodactyls or cats (and for that matter, any other non-primate mammal). In a mammal with fewer A3 genes, A3 proteins may be more ubiquitously expressed, like in the mouse (36), whereas in a mammal with a larger number of genes, expression may be more specific to certain tissues reflecting functional specialization. For example, human A3A plays a central role in restricting foreign DNA from a variety of intracellular and extracellular sources and is expressed predominantly (if not exclusively) in phagocytic cells (135, 160).

As the primary function of A3 proteins is to restrict exogenous retroviruses and endogenous retroelements, and it is likely that selective pressure from one or both of these sources drove expansion of the locus in select lineages. A comparative analysis of the human and mouse genomes show that transposition activity in the mouse has been consistently active for the last 100 mya whereas in the human, transposition activity has dropped off in the last 40 mya (85, 176). It is possible that transposition activity was the pressure to select for a larger A3 repertoire, protecting our common primate ancestor from lethal insertions and preserving genome integrity. On the other hand, A3Z2-Z3 knockout mice have been created and were found to be more susceptible to infection with mouse leukemia virus (MLV) (145, 165) and mouse mammary tumor virus (MMTV)

(127). Thus, it is equally as possible that a larger A3 repertoire has been selected in response to an early retroviral pathogen. The advantage of having a larger A3 repertoire may become clearer as more mammalian genomes are completely sequenced and more knockout studies are done in cell lines from different mammals. A correlation may be revealed between mammals containing different numbers of A3 genes and also may explain a possible functional advantage between the different A3 Z-types.

Mammalian A3Z3 proteins are neutralized by their species specific Vif

In Chapter 3, the mammalian single domain Z3-type A3 proteins of human, rhesus macaque, cow, sheep and cat were utilized to investigate if the mammalian A3/lentiviral Vif interaction is conserved. All of these mammalian A3Z3 proteins are all able to restrict HIV to varying degrees and are found to be neutralized in the presence of their species specific Vif. Moreover, Vifs from SIVmac and MVV are able to neutralize other species A3Z3 proteins. The data collected from these comparative experiments are supportive of a conserved, interaction between mammalian A3Z3 proteins and lentiviral Vif proteins. The major unanswered question is at what level is this interaction conserved (**Figure 4-3**)?

Vif neutralization of A3Z3 proteins: proteasome-dependent or -independent?

Host A3Z3 proteins appear to be degraded in the presence of their species lentiviral Vif [Chapter 3 (89)]. As reviewed in Chapter 1, HIV-1 Vif binds to human A3G and recruits a CULLIN-5 based E3 ubiquitin ligase complex which results in polyubiquitination and degradation of A3G. The experiments in Chapter 3 were designed

to test for a proteasome-dependent pathway by mutating the highly conserved S/TLQ motif in each lentiviral Vif, which in HIV Vif is essential for binding of the ELONGINB/C heterodimer (180). The data show that in the presence of each lentiviral Vif S/TLQ mutant there was very little or no degradation detected in all of the A3Z3 proteins, supporting the possibility that all lentiviral Vif proteins act as an adaptor protein to a CULLIN 5-based E3 ubiquitin ligase. However it is important to note that primate lentiviral Vif proteins sequences share a conserved zinc-binding motif ($Hx_5Cx_2Cx_3H$) not found in BIV, MVV or FIV Vif sequences (181). So it is possible that non-primate lentiviral Vif proteins utilize a CULLIN-2 based E3 ubiqutin ligase that consists of the ELONGINB/C heterodimer and the RBX2 proteins. There are four other CULLIN proteins (1, 3, 4A, 4B) that rely on an adaptor protein (example Vif) to bind to their target substrate (example A3G) that do not utilize the ELONGINB/C heterodimer (reviewed in (65)). The SIVmac Vif protein may have evolved to hijack another CULLIN-based E3 ubiquitin ligase and could explain why in the presence of the SIVmac Vif SLQ mutant protein, some level of degradation in human and rhesus A3H proteins is observed. The next logical step is to confirm an interaction between each A3Z3 and Vif protein and determine the interacting protein components of the CULLIN-based E3 ubiquitin ligase complex.

Alternatively, there is evidence that Vif may sometimes be utilizing a proteasome-independent pathway. In these instances, such as cow and sheep A3Z3 in the presence of HIV Vif (Chapter 3 **Figure 3-4**), a rescue of HIV infectivity is observed without concomitant A3Z3 degradation by Vif. Again, determining an interaction between each protein is important but further experimentation is required to determine how Vif is not

serving as an adaptor protein for cellular proteasome degradation. The ability of certain Vifs to see multiply species A3Z3 proteins may help explain this observation. A 'promiscuous' Vif may see a conserved region and/or structural element on all A3 proteins that makes it 'sticky' but not able to efficiently recruit or place an E3 ubiquitin ligase complex. It may bind in a way that would cause a protein conformational change or steric hindrance, interfering with the A3 proteins ability to restrict virus.

The Vif-interacting region of A3

As discussed in Chapter 3, except for the highly conserved core residues, the overall sequence identity and similarity between the predicted surface amino acid residues of mammalian A3 proteins is low. In previous studies, patches of these differing residues were swapped between primate species to identify the HIV Vif-interacting regions in human A3G (11, 103, 104, 142, 147, 182) and A3F (142). More recently, swapping residues between different human A3H haplotypes with varying HIV-1 Vif sensitivities identified a similar putative Vif-interacting region(31, 95). When these Z2 and Z3 –type A3 sequences are aligned, the identifed Vif –sensitive residues all fall into a candidate Vif-interacting region of approximately 34 residues (**Figure 4-4**). Adding the high-resolution A3Gctd structures (20, 43, 54, 62, 152) to the predicted corresponding regions of the human A3F, A3G and A3H alignment maps this putative Vif-interacting region to the alpha3-loop-beta4-loop-alpha4 (Albin, LaRue et al., in preparation).

This region of Vif-sensitive residues in different human A3 proteins, along with predicted secondary structure helps provide a starting point for investigating the Vif-sensitive regions of other mammalian A3Z3 proteins. Experiments involving A3Z3

chimeras can be designed to determine the Vif-sensitive regions for each species A3Z3 protein. The SIVmac and MVV Vif proteins will be particularly useful in delineating a conserved Vif-interacting region in mammalian A3 proteins because of their ability to see the A3Z3 proteins from multiple species as well as other Z2-type A3 containing proteins [Chapter 3 (11, 89, 147, 174, 189)]. This experimental approach is versatile and could be used to test the entire matrix of lentiviral Vifs with all mammalian A3 proteins in order to map the Vif-sensitive regions.

It is plausible (but non-parsimonious) that the Vif-interacting region of other mammalian A3 proteins is completely different and varies between species. This is possible given the immensely high degree of sequence divergence between different lentiviral Vif proteins. A conserved functional interaction despite a divergent interacting region would highlight the versatility of Vif and lentiviruses in general in evading host defenses. Either outcome, conserved or not, will provide substantial insight into the design of efficient therapeutics to inhibit Vif in lentivirus-induced diseases. A couple of small molecule compounds have been identified that prevent the replication of Vif-proficient HIV in A3G-expressing cells. One of the molecules, RN-18, enhances Vif degradation only in the presence of A3G, suggesting a direct interaction with Vif (119). The other two molecules identified, IMB-26 and 35, appear to bind to human A3G directly, disrupting the binding of HIV Vif (17). Testing these difference compounds for their ability to disrupt other mammalian A3/Vif protein interactions may provide evidence of how robust each inhibitor is and if the Vif-interacting region is a relevant target in ameliorating a broader array of lentiviral diseases.

The A3-interacting region of Vif

Numerous studies have shown that multiple regions within the NTD of HIV Vif are involved in binding the human A3 proteins (18, 31, 60, 109, 131, 141, 148, 155, 169, 184, 193) (reviewed in Chapter 1, **Figure 1-5**) and SIVmac Vif (18, 31). To date, there is no high-resolution structure of the HIV Vif protein with or without an interacting A3 protein. Therefore, it is unknown how changing the A3-sensitive residues identified on HIV-1 Vif may affect the integrity of Vif's overall structural conformation and surface charge. A3H is the only A3 protein in the human repertoire whose general interacting region on Vif has not yet been identified. This is due in part because a stable A3H haplotype has only recently been identified (125) and its interaction with Vif is weak (31, 89, 95) or undetectable (53) (reported in Chapter 3).

Determining the A3-interacting region of all lentiviral Vifs used in this study is the more challenging component of this hypothesis. Sequence identity is very low between lentiviral Vif proteins, especially in the NTD between primate and non-primate Vif proteins where human A3 interacting regions are located in HIV-1 Vif (Chapter 1, **Figure 1-5**). The solution to this problem will probably require high-resolution structures for all of the Vif proteins to be able to identify conserved structural motifs.

Refining the working hypothesis

The work presented in this dissertation provides a substantial foundation for determining at what level the mammalian A3/ lentiviral Vif interaction is conserved. We can now start to refine our hypothesis outlined in **Figure 4-3**. This schematic highlights the levels of the A3/Vif relationship that are conserved and not conserved. We were able

to immediately rule out a conserved sequence motif that would be important for the A3/Vif interaction on both proteins. This could be predicted due to the observation that proteins involved in a host-pathogen interaction tend to be under positive selection (thus low sequence identity). On the other end of the spectrum of conservation, we show that there is a relationship or 'functional interplay' between host A3Z3s and their species lentiviral Vifs (aka degradation of A3Z3 protein). Furthermore, the lack of degradation in the presence of each mutant Vif (S/TLQ) implies that a proteasome-dependent degradation mechanism may be conserved between different species, but more experiments are needed to confirm that E3 ligase complex components are present. What remains unknown is whether a conserved interacting region is present in all mammalian host A3s and lentiviral Vifs. So the refined hypothesis includes three major facets: (1) Lentiviral Vif proteins utilize a proteasome-dependent Cullin-based E3 ubiquitin ligase to target mammalian A3s for degradation, (2) a Vif-interacting site spans a conserved region on all Vif-sensitive A3 proteins, and (3) the A3-interacting site of different lentiviral Vif proteins is not conserved and may be more diffuse.

Five years ago, our perception was limited because it was not known how many A3 proteins existed in these mammals and if a relationship with their species lentiviral Vif even existed. A comparative approach to this particular host-pathogen conflict, highlights the complexity of how host proteins and viral proteins adapt and evolve to counteract each other. The findings of this research could provide valuable insight in to the design of therapeutics to effectively target proteins like Vif and allow A3 proteins to restrict HIV.

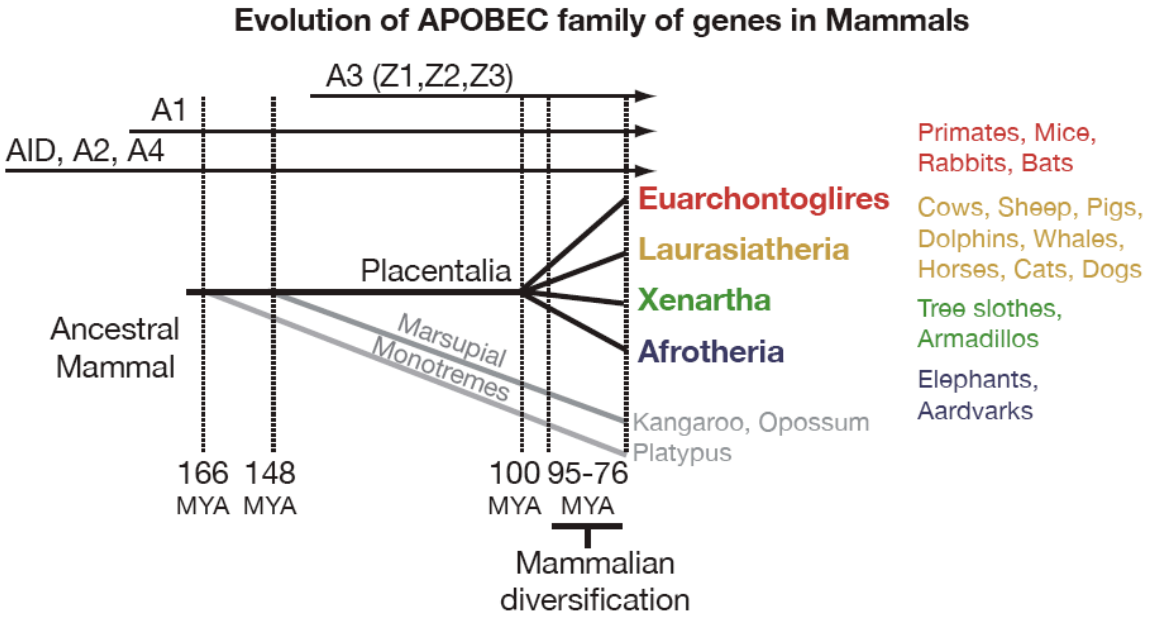


Figure 4-1. Schematic of mammalian evolution and the appearance of A3 genes. The first number listed at the left-hand bottom indicate the approximate time in millions of years since the divergence of Monotremes from the common ancestor to Marsupials and Placentalia and the second number the divergence of Marupials from Placentalia based on calculations in (9). The third number from the left indicates the split into the mammalian super clades (shown in bold colors), the range of numbers indicate a time of rapid diversification in all placentalia super-clades. Species on the right are color-coded to match with their super clade. The black arrow on top is meant to correlate with the presence of APOBEC family of genes and mammals.

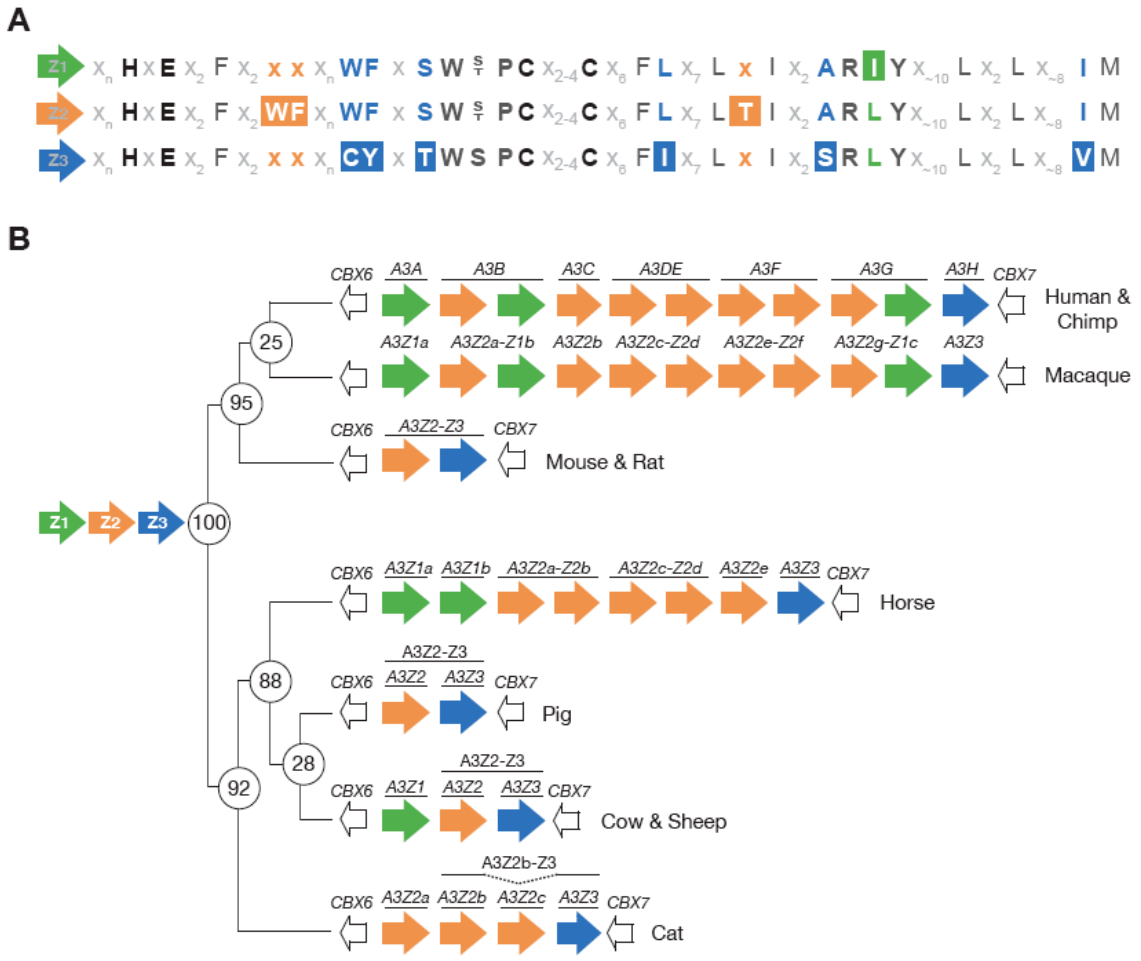


Figure 4-2. (A) Highlights of the amino acid conservation among the three distinct Z-domain groups and within each individual group [based on multiple sequence alignments (88)]. Residues discussed in the text are colored or emboldened, and other notable residues are shaded grey. An 'X' specifies nearly any amino acid.

(B) Schematics of the *A3* repertoires of mammals whose genomes have been sequenced. Z1, Z2, and Z3 domains are colored in green, orange and blue, respectively. For all of the indicated species (and likely all mammals), *CBX6* is located immediately upstream and *CBX7* downstream of the *A3* locus. The inferred ancestral *A3* repertoire was deduced through comparative studies (88). The numbers at the phylogenetic tree 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 (9). Adapted from LaRue et al., 2009 (Appendix 1) (87).

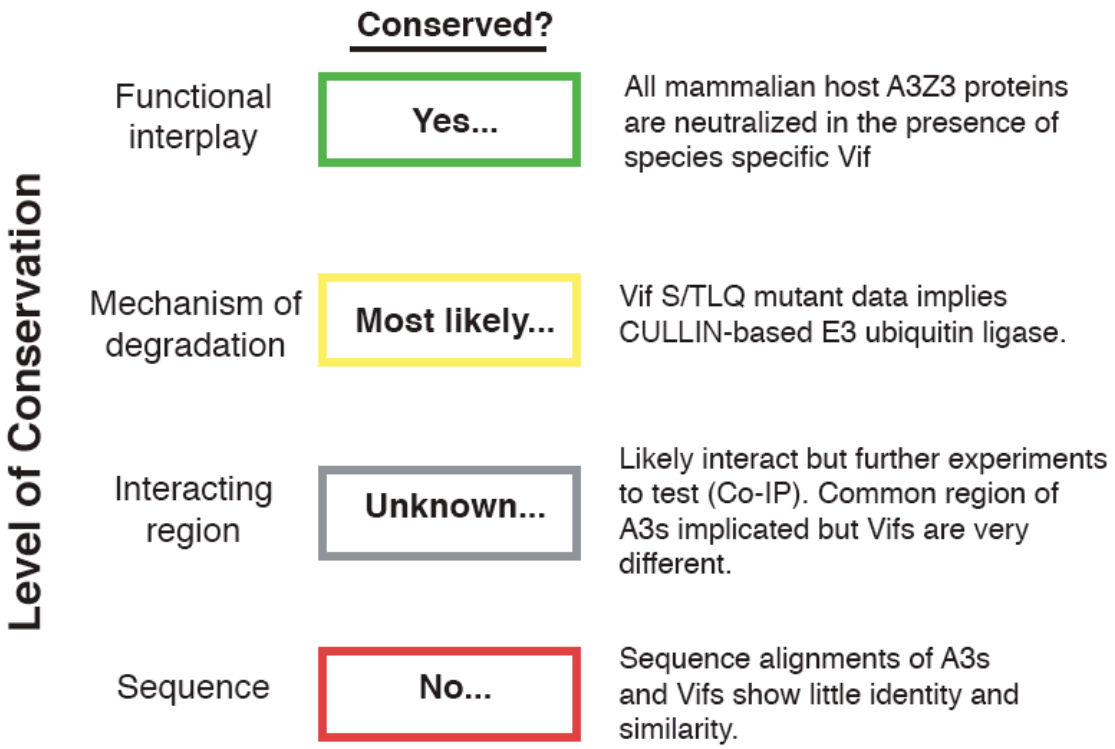


Figure 4-3. A schematic depicting the different levels of conservation for the mammalian A3/Lentiviral Vif interaction.

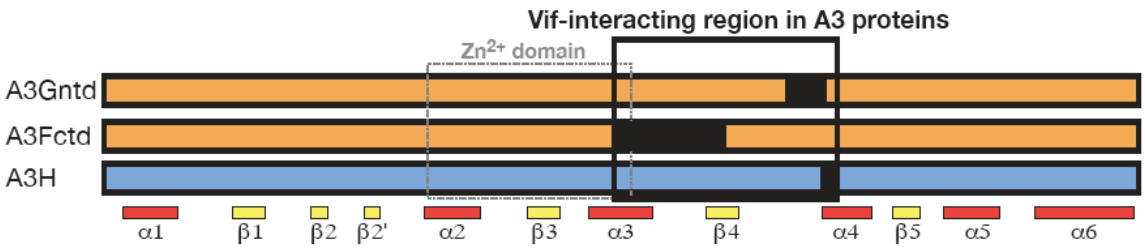


Figure 4-4. A span of HIV-1 Vif-susceptible A3 residues in human A3 proteins. The solid black boxes indicate the region sensitive to HIV-1 Vif for each individual A3 domain and the open box indicates the Vif sensitive region (references in text). The secondary structure based on the high-resolution structure for human A3Gctd (20, 43, 54, 62, 152) is placed according to an alignment of A3Gctd, A3Gntd, A3Fctd and A3H domains. The orange color represents the Z2-type A3 and the blue color represents the Z3-type A3 (based on figure 4-1).

Bibliography

1. Abrahams, V. M., T. M. Schaefer, J. V. Fahey, I. Visintin, J. A. Wright, P. B. Aldo, R. Romero, C. R. Wira, and G. Mor. 2006. Expression and secretion of antiviral factors by trophoblast cells following stimulation by the TLR-3 agonist, Poly(I : C). Hum Reprod 21:2432-9.
2. Abudu, A., A. Takaori-Kondo, T. Izumi, K. Shirakawa, M. Kobayashi, A. Sasada, K. Fukunaga, and T. Uchiyama. 2006. Murine retrovirus escapes from murine APOBEC3 via two distinct novel mechanisms. Curr Biol 16:1565-70.
3. Albin, J. S., and R. S. Harris. Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics. Expert Rev Mol Med 12:e4.
4. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J Mol Biol 215:403-10.
5. Anant, S., and N. O. Davidson. 2001. Molecular mechanisms of apolipoprotein B mRNA editing. Curr Opin Lipidol 12:159-65.
6. Armitage, A. E., A. Katzourakis, T. de Oliveira, J. J. Welch, R. Belshaw, K. N. Bishop, B. Kramer, A. J. McMichael, A. Rambaut, and A. K. Iversen. 2008. Conserved footprints of APOBEC3G on Hypermutated human immunodeficiency virus type 1 and human endogenous retrovirus HERV-K(HML2) sequences. J Virol 82:8743-61.
7. Bertrand, D., M. Lajoie, and N. El-Mabrouk. 2008. Inferring ancestral gene orders for a family of tandemly arrayed genes. J Comput Biol 15:1063-77.
8. Bieniasz, P. D. 2004. Intrinsic immunity: a front-line defense against viral attack. Nat Immunol 5:1109-15.
9. Bininda-Emonds, O. R., M. Cardillo, K. E. Jones, R. D. MacPhee, R. M. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis. 2007. The delayed rise of present-day mammals. Nature 446:507-12.
10. Bishop, K. N., R. K. Holmes, A. M. Sheehy, N. O. Davidson, S. J. Cho, and M. H. Malim. 2004. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Curr Biol 14:1392-6.
11. Bogerd, H. P., B. P. Doehle, H. L. Wiegand, and B. R. Cullen. 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 U S A 101:3770-4.
12. Bogerd, H. P., R. L. Tallmadge, J. L. Oaks, S. Carpenter, and B. R. Cullen. 2008. Equine infectious anemia virus resists the antiretroviral activity of equine APOBEC3 proteins through a packaging-independent mechanism. J Virol 82:11889-901.
13. Bogerd, H. P., H. L. Wiegand, B. P. Doehle, and B. R. Cullen. 2007. The intrinsic antiretroviral factor APOBEC3B contains two enzymatically active cytidine deaminase domains. Virology 364:486-93.
14. Bogerd, H. P., H. L. Wiegand, A. E. Hulme, J. L. Garcia-Perez, K. S. O'Shea, J. V. Moran, and B. R. Cullen. 2006. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci U S A 103:8780-5.
15. Bonvin, M., F. Achermann, I. Greeve, D. Stroka, A. Keogh, D. Inderbitzin, D. Candinas, P. Sommer, S. Wain-Hobson, J. P. Vartanian, and J. Greeve. 2006. Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. Hepatology 43:1364-74.
16. Brodie, R., R. L. Roper, and C. Upton. 2004. JDotter: a Java interface to multiple dotplots generated by dotter. Bioinformatics 20:279-81.
17. Cen, S., Z. G. Peng, X. Y. Li, Z. R. Li, J. Ma, Y. M. Wang, B. Fan, X. F. You, Y. P. Wang, F. Liu, R. G. Shao, L. X. Zhao, L. Yu, and J. D. Jiang. 2010. Small molecular compounds inhibit HIV-1 replication through specifically stabilizing APOBEC3G. J Biol Chem 285:16546-52.

18. Chen, G., Z. He, T. Wang, R. Xu, and X. F. Yu. 2009. A patch of positively charged amino acids surrounding the human immunodeficiency virus type 1 Vif SLVx4Yx9Y motif influences its interaction with APOBEC3G. J Virol 83:8674-82.
19. Chen, H., C. E. Lilley, Q. Yu, D. V. Lee, J. Chou, I. Narvaiza, N. R. Landau, and M. D. Weitzman. 2006. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. Curr Biol 16:480-5.
20. Chen, K. M., E. Harjes, P. J. Gross, A. Fahmy, Y. Lu, K. Shindo, R. S. Harris, and H. Matsuo. 2008. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. Nature 452:116-9.
21. Chen, K. M., N. Martemyanova, Y. Lu, K. Shindo, H. Matsuo, and R. S. Harris. 2007. Extensive mutagenesis experiments corroborate a structural model for the DNA deaminase domain of APOBEC3G. FEBS Lett 581:4761-6.
22. Chiu, Y. L., and W. C. Greene. 2008. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annu Rev Immunol 26:317-53.
23. Clamp, M., J. Cuff, S. M. Searle, and G. J. Barton. 2004. The Jalview Java alignment editor. Bioinformatics 20:426-7.
24. Clements, J. E., and M. C. Zink. 1996. Molecular biology and pathogenesis of animal lentivirus infections. Clin Microbiol Rev 9:100-17.
25. Conticello, S. G., R. S. Harris, and M. S. Neuberger. 2003. The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. Curr Biol 13:2009-13.
26. Conticello, S. G., C. J. Thomas, S. K. Petersen-Mahrt, and M. S. Neuberger. 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol 22:367-77.
27. Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner. 2004. WebLogo: a sequence logo generator. Genome Res 14:1188-90.
28. Cullen, B. R. 2006. Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors. J Virol 80:1067-76.
29. Dang, Y., L. M. Siew, X. Wang, Y. Han, R. Lampen, and Y. H. Zheng. 2008. Human cytidine deaminase APOBEC3H restricts HIV-1 replication. J Biol Chem 283:11606-14.
30. Dang, Y., X. Wang, W. J. Esselman, and Y. H. Zheng. 2006. Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. J Virol 80:10522-33.
31. Dang, Y., X. Wang, T. Zhou, I. A. York, and Y. H. Zheng. 2009. Identification of a novel WxSLVK motif in the N terminus of human immunodeficiency virus and simian immunodeficiency virus Vif that is critical for APOBEC3G and APOBEC3F neutralization. J Virol 83:8544-52.
32. Delebecque, F., R. Suspene, S. Calattini, N. Casartelli, A. Saib, A. Froment, S. Wain-Hobson, A. Gessain, J. P. Vartanian, and O. Schwartz. 2006. Restriction of foamy viruses by APOBEC cytidine deaminases. J Virol 80:605-14.
33. Derse, D., S. A. Hill, G. Princler, P. Lloyd, and G. Heidecker. 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 104:2915-20.
34. 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 72:1431-7.
35. Doehle, B. P., H. P. Bogerd, H. L. Wiegand, N. Jouvenet, P. D. Bieniasz, E. Hunter, and B. R. Cullen. 2006. The betaretrovirus Mason-Pfizer monkey virus selectively excludes simian APOBEC3G from virion particles. J Virol 80:12102-8.
36. Doehle, B. P., A. Schafer, H. L. Wiegand, H. P. Bogerd, and B. R. Cullen. 2005. Differential sensitivity of murine leukemia virus to APOBEC3-mediated inhibition is governed by virion exclusion. J Virol 79:8201-7.
37. Dutko, J. A., A. Schafer, A. E. Kenny, B. R. Cullen, and M. J. Curcio. 2005. Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases. Curr Biol 15:661-6.

38. Esnault, C., O. Heidmann, F. Delebecque, M. Dewannieux, D. Ribet, A. J. Hance, T. Heidmann, and O. Schwartz. 2005. APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. Nature 433:430-3.
39. Eto, T., K. Kinoshita, K. Yoshikawa, M. Muramatsu, and T. Honjo. 2003. RNA-editing cytidine deaminase Apobec-1 is unable to induce somatic hypermutation in mammalian cells. Proc Natl Acad Sci U S A 100:12895-8.
40. Fahrenkrug, S. C., G. A. Rohrer, B. A. Freking, T. P. Smith, K. Osoegawa, C. L. Shu, J. J. Catanese, and P. J. de Jong. 2001. A porcine BAC library with tenfold genome coverage: a resource for physical and genetic map integration. Mamm Genome 12:472-4.
41. Fahrenkrug, S. C., T. P. Smith, B. A. Freking, J. Cho, J. White, J. Vallet, T. Wise, G. Rohrer, G. Pertea, R. Sultana, J. Quackenbush, and J. W. Keele. 2002. Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly. Mamm Genome 13:475-8.
42. Fitch, W. M. 1977. 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 86:623-44.
43. Furukawa, A., T. Nagata, A. Matsugami, Y. Habu, R. Sugiyama, F. Hayashi, N. Kobayashi, S. Yokoyama, H. Takaku, and M. Katahira. 2009. Structure and real-time monitoring of the enzymatic reaction of APOBEC3G which is involved in anti-HIV activity. Nucleic Acids Symp Ser (Oxf):87-8.
44. Gabuzda, D. H., A. Lever, E. Terwilliger, and J. Sodroski. 1992. Effects of deletions in the cytoplasmic domain on biological functions of human immunodeficiency virus type 1 envelope glycoproteins. J Virol 66:3306-15.
45. Gabuzda, D. H., H. Li, K. Lawrence, B. S. Vasir, K. Crawford, and E. Langhoff. 1994. Essential role of vif in establishing productive HIV-1 infection in peripheral blood T lymphocytes and monocyte/macrophages. J Acquir Immune Defic Syndr 7:908-15.
46. Gascuel, O., D. Bertrand, O. Elemento. 2007. Reconstructing the duplication history of tandemly repeated sequences, p. 205-235, Mathematics of Evolution and Phylogeny.
47. Gibbs, J. S., D. A. Regier, and R. C. Desrosiers. 1994. Construction and in vitro properties of HIV-1 mutants with deletions in "nonessential" genes. AIDS Res Hum Retroviruses 10:343-50.
48. Gibbs, R. A., J. Rogers, M. G. Katze, R. Bumgarner, G. M. Weinstock, E. R. Mardis, K. A. Remington, R. L. Strausberg, J. C. Venter, R. K. Wilson, M. A. Batzer, C. D. Bustamante, E. E. Eichler, M. W. Hahn, R. C. Hardison, K. D. Makova, W. Miller, A. Milosavljevic, R. E. Palermo, A. Siepel, J. M. Sikela, T. Attaway, S. Bell, K. E. Bernard, C. J. Buhay, M. N. Chandrabose, M. Dao, C. Davis, K. D. Delehaunty, Y. Ding, H. H. Dinh, S. Dugan-Rocha, L. A. Fulton, R. A. Gabisi, T. T. Garner, J. Godfrey, A. C. Hawes, J. Hernandez, S. Hines, M. Holder, J. Hume, S. N. Jhangiani, V. Joshi, Z. M. Khan, E. F. Kirkness, A. Cree, R. G. Fowler, S. Lee, L. R. Lewis, Z. Li, Y. S. Liu, S. M. Moore, D. Muzny, L. V. Nazareth, D. N. Ngo, G. O. Okwuonu, G. Pai, D. Parker, H. A. Paul, C. Pfannkoch, C. S. Pohl, Y. H. Rogers, S. J. Ruiz, A. Sabo, J. Santibanez, B. W. Schneider, S. M. Smith, E. Sodergren, A. F. Svatek, T. R. Utterback, S. Vattathil, W. Warren, C. S. White, A. T. Chinwalla, Y. Feng, A. L. Halpern, L. W. Hillier, X. Huang, P. Minx, J. O. Nelson, K. H. Pepin, X. Qin, G. G. Sutton, E. Venter, B. P. Walenz, J. W. Wallis, K. C. Worley, S. P. Yang, S. M. Jones, M. A. Marra, M. Rocchi, J. E. Schein, R. Baertsch, L. Clarke, M. Csuros, J. Glasscock, R. A. Harris, P. Havlak, A. R. Jackson, H. Jiang, et al. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. Science 316:222-34.
49. Gifford, R. J., A. Katzourakis, M. Tristem, O. G. Pybus, M. Winters, and R. W. Shafer. 2008. A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. Proc Natl Acad Sci U S A 105:20362-7.
50. Goff, S. P. 2004. Retrovirus restriction factors. Mol Cell 16:849-59.
51. Goila-Gaur, R., and K. Strebel. 2008. HIV-1 Vif, APOBEC, and intrinsic immunity. Retrovirology 5:51.

52. Haché, G., M. T. Liddament, and R. S. Harris. 2005. The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. J Biol Chem 280:10920-4.
53. Harari, A., M. Ooms, L. C. Mulder, and V. Simon. 2009. Polymorphisms and splice variants influence the antiretroviral activity of human APOBEC3H. J Virol 83:295-303.
54. Harjes, E., P. J. Gross, K. M. Chen, Y. Lu, K. Shindo, R. Nowarski, J. D. Gross, M. Kotler, R. S. Harris, and H. Matsuo. 2009. An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. J Mol Biol 389:819-32.
55. 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 69:3247-57.
56. 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 113:803-9.
57. Harris, R. S., and M. T. Liddament. 2004. Retroviral restriction by APOBEC proteins. Nat Rev Immunol 4:868-77.
58. Harris, R. S., S. K. Petersen-Mahrt, and M. S. Neuberger. 2002. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. Mol Cell 10:1247-53.
59. Harris, R. S., R. S. LaRue, T. J. Hope. In press. Retrovirus Restriction Factors. *In* A. Lever, B, Berkhout, K. Jeang (ed.), Human Retroviruses. World Scientific Publishing, Cambridge, Amsterdam, Bethesda.
60. He, Z., W. Zhang, G. Chen, R. Xu, and X. F. Yu. 2008. Characterization of conserved motifs in HIV-1 Vif required for APOBEC3G and APOBEC3F interaction. J Mol Biol 381:1000-11.
61. Henikoff, S., and J. G. Henikoff. 1992. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 89:10915-9.
62. Holden, L. G., C. Prochnow, Y. P. Chang, R. Bransteitter, L. Chelico, U. Sen, R. C. Stevens, M. F. Goodman, and X. S. Chen. 2008. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. Nature 456:121-4.
63. Hultquist, J. F., and R. S. Harris. 2009. Leveraging APOBEC3 proteins to alter the HIV mutation rate and combat AIDS. Future Virol 4:605.
64. Iwatani, Y., H. Takeuchi, K. Strebel, and J. G. Levin. 2006. Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. J Virol 80:5992-6002.
65. Jackson, S., and Y. Xiong. 2009. CRL4s: the CUL4-RING E3 ubiquitin ligases. Trends Biochem Sci 34:562-70.
66. Jarmuz, A., A. Chester, J. Bayliss, J. Gisbourne, I. Dunham, J. Scott, and N. Navaratnam. 2002. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. Genomics 79:285-96.
67. Jern, P., J. P. Stoye, and J. M. Coffin. 2007. Role of APOBEC3 in genetic diversity among endogenous murine leukemia viruses. PLoS Genet 3:2014-22.
68. Jónsson, S. R., G. Haché, M. D. Stenglein, S. C. Fahrenkrug, V. Andrésdóttir, and R. S. Harris. 2006. Evolutionarily conserved and non-conserved retrovirus restriction activities of artiodactyl APOBEC3F proteins. Nucleic Acids Res 34:5683-94.
69. Jonsson, S. R., R. S. LaRue, M. D. Stenglein, S. C. Fahrenkrug, V. Andresdottir, and R. S. Harris. 2007. The restriction of zoonotic PERV transmission by human APOBEC3G. PLoS ONE 2:e893.
70. Jost, S., P. Turelli, B. Mangeat, U. Protzer, and D. Trono. 2007. Induction of antiviral cytidine deaminases does not explain the inhibition of hepatitis B virus replication by interferons. J Virol 81:10588-96.
71. Kaiser, S. M., H. S. Malik, and M. Emerman. 2007. Restriction of an extinct retrovirus by the human TRIM5alpha antiviral protein. Science 316:1756-8.
72. Kao, S., R. Goila-Gaur, E. Miyagi, M. A. Khan, S. Opi, H. Takeuchi, and K. Strebel. 2007. Production of infectious virus and degradation of APOBEC3G are separable functional properties of human immunodeficiency virus type 1 Vif. Virology 369:329-39.

73. Katzourakis, A., M. Tristem, O. G. Pybus, and R. J. Gifford. 2007. Discovery and analysis of the first endogenous lentivirus. Proc Natl Acad Sci U S A 104:6261-5.
74. Kessler, D. S., D. E. Levy, and J. E. Darnell, Jr. 1988. Two interferon-induced nuclear factors bind a single promoter element in interferon-stimulated genes. Proc Natl Acad Sci U S A 85:8521-5.
75. Kidd, J. M., T. L. Newman, E. Tuzun, R. Kaul, and E. E. Eichler. 2007. Population stratification of a common APOBEC gene deletion polymorphism. PLoS Genet 3:e63.
76. Kinomoto, M., T. Kanno, M. Shimura, Y. Ishizaka, A. Kojima, T. Kurata, T. Sata, and K. Tokunaga. 2007. All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. Nucleic Acids Res 35:2955-64.
77. Kobayashi, M., A. Takaori-Kondo, Y. Miyauchi, K. Iwai, and T. Uchiyama. 2005. Ubiquitination of APOBEC3G by an HIV-1 Vif-Cullin5-Elongin B-Elongin C complex is essential for Vif function. J Biol Chem 280:18573-8.
78. Kobayashi, M., A. Takaori-Kondo, K. Shindo, A. Abudu, K. Fukunaga, and T. Uchiyama. 2004. APOBEC3G targets specific virus species. J Virol 78:8238-44.
79. Koning, F. A., E. N. Newman, E. Y. Kim, K. J. Kunstman, S. M. Wolinsky, and M. H. Malim. 2009. Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. J Virol 83:9474-85.
80. Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. Frost. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23:1891-901.
81. Kristbjornsdottir, H. B., V. Andresdottir, V. Svansson, S. Torsteinsdottir, S. Matthiasdottir, and O. S. Andresson. 2004. The vif gene of maedi-visna virus is essential for infectivity in vivo and in vitro. Virology 318:350-9.
82. Krumsiek, J., R. Arnold, and T. Rattei. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics 23:1026-8.
83. Kumar, S., and S. B. Hedges. 1998. A molecular timescale for vertebrate evolution. Nature 392:917-20.
84. Lajoie, M., D. Bertrand, N. El-Mabrouk, and O. Gascuel. 2007. Duplication and inversion history of a tandemly repeated genes family. J Comput Biol 14:462-78.
85. Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860-921.
86. Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-8.
87. LaRue, R. S., V. Andrésdóttir, Y. Blanchard, S. G. Conticello, D. Derse, M. Emerman, W. C. Greene, S. R. Jónsson, N. R. Landau, M. Löchelt, H. S. Malik, M. H. Malim, C. Münk, S. J. O'Brien, V. K. Pathak, K. Strebel, S. Wain-Hobson, X. F. Yu, N. Yuhki, and R. S. Harris. 2009. Guidelines for naming nonprimate APOBEC3 genes and proteins. J Virol 83:494-7.
88. LaRue, R. S., S. R. Jónsson, K. A. Silverstein, M. Lajoie, D. Bertrand, N. El-Mabrouk, I. Hötzel, V. Andrésdóttir, T. P. Smith, and R. 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 Mol Biol 9:104.
89. Larue, R. S., J. Lengyel, S. R. Jonsson, V. Andresdottir, and R. S. Harris. 2010. Lentiviral Vif Degrades the APOBEC3Z3 Protein of its Mammalian Host and is Capable of Cross-Species Activity. J Virol.
90. 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.
91. Lee, Y. N., and P. D. Bieniasz. 2007. Reconstitution of an infectious human endogenous retrovirus. PLoS Pathog 3:e10.
92. Lee, Y. N., M. H. Malim, and P. D. Bieniasz. 2008. Hypermutation of an ancient human retrovirus by APOBEC3G. J Virol 82:8762-70.
93. Leroux, C., J. L. Cadore, and R. C. Montelaro. 2004. Equine Infectious Anemia Virus (EIAV): what has HIV's country cousin got to tell us? Vet Res 35:485-512.
94. Levy, D. E., D. S. Kessler, R. Pine, N. Reich, and J. E. Darnell, Jr. 1988. Interferon-induced nuclear factors that bind a shared promoter element correlate with positive and negative transcriptional control. Genes Dev 2:383-93.
95. Li, M. M., L. I. Wu, and M. Emerman. The range of human APOBEC3H sensitivity to lentiviral Vif proteins. J Virol 84:88-95.
96. Liao, W., S. H. Hong, B. H. Chan, F. B. Rudolph, S. C. Clark, and L. Chan. 1999. APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. Biochem Biophys Res Commun 260:398-404.
97. Liddament, M. T., W. L. Brown, A. J. Schumacher, and R. S. Harris. 2004. APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. Curr Biol 14:1385-91.
98. Lochelt, M., F. Romen, P. Bastone, H. Muckenfuss, N. Kirchner, Y. B. Kim, U. Truyen, U. Rosler, M. Battenberg, A. Saib, E. Flory, K. Cichutek, and C. Munk. 2005. The antiretroviral activity of APOBEC3 is inhibited by the foamy virus accessory Bet protein. Proc Natl Acad Sci U S A 102:7982-7.
99. 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.
100. Luo, K., Z. Xiao, E. Ehrlich, Y. Yu, B. Liu, S. Zheng, and X. F. Yu. 2005. Primate lentiviral virion infectivity factors are substrate receptors that assemble with cullin 5-E3 ligase through a HCCH motif to suppress APOBEC3G. Proc Natl Acad Sci U S A 102:11444-9.
101. Malim, M. H., and M. Emerman. 2008. HIV-1 accessory proteins--ensuring viral survival in a hostile environment. Cell Host Microbe 3:388-98.
102. 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.
103. Mangeat, B., P. Turelli, S. Liao, and D. Trono. 2004. A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action. J Biol Chem 279:14481-3.
104. Mariani, R., D. Chen, B. Schröfelbauer, F. Navarro, R. König, B. Bollman, C. Münk, H. Nymark-McMahon, and N. R. Landau. 2003. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. Cell 114:21-31.
105. Marin, M., S. Golem, K. M. Rose, S. L. Kozak, and D. Kabat. 2008. Human immunodeficiency virus type 1 Vif functionally interacts with diverse APOBEC3 cytidine deaminases and moves with them between cytoplasmic sites of mRNA metabolism. J Virol 82:987-98.
106. Marin, M., K. M. Rose, S. L. Kozak, and D. Kabat. 2003. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. Nat Med 9:1398-403.
107. Matsumoto, T., H. Marusawa, Y. Endo, Y. Ueda, Y. Matsumoto, and T. Chiba. 2006. Expression of APOBEC2 is transcriptionally regulated by NF-kappaB in human hepatocytes. FEBS Lett 580:731-5.

108. Mehle, A., B. Strack, P. Ancuta, C. Zhang, M. McPike, and D. Gabuzda. 2004. Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. J Biol Chem 279:7792-8.

109. Mehle, A., H. Wilson, C. Zhang, A. J. Brazier, M. McPike, E. Pery, and D. Gabuzda. 2007. Identification of an APOBEC3G binding site in human immunodeficiency virus type 1 Vif and inhibitors of Vif-APOBEC3G binding. J Virol 81:13235-41.

110. Mikkelsen, T. S., M. J. Wakefield, B. Aken, C. T. Amemiya, J. L. Chang, S. Duke, M. Garber, A. J. Gentles, L. Goodstadt, A. Heger, J. Jurka, M. Kamal, E. Mauceli, S. M. Searle, T. Sharpe, M. L. Baker, M. A. Batzer, P. V. Benos, K. Belov, M. Clamp, A. Cook, J. Cuff, R. Das, L. Davidow, J. E. Deakin, M. J. Fazzari, J. L. Glass, M. Grabherr, J. M. Greally, W. Gu, T. A. Hore, G. A. Huttley, M. Kleber, R. L. Jirtle, E. Koina, J. T. Lee, S. Mahony, M. A. Marra, R. D. Miller, R. D. Nicholls, M. Oda, A. T. Papenfuss, Z. E. Parra, D. D. Pollock, D. A. Ray, J. E. Schein, T. P. Speed, K. Thompson, J. L. VandeBerg, C. M. Wade, J. A. Walker, P. D. Waters, C. Webber, J. R. Weidman, X. Xie, M. C. Zody, J. A. Graves, C. P. Ponting, M. Breen, P. B. Samollow, E. S. Lander, and K. Lindblad-Toh. 2007. Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. Nature 447:167-77.

111. Miyoshi, H., U. Blomer, M. Takahashi, F. H. Gage, and I. M. Verma. 1998. Development of a self-inactivating lentivirus vector. J Virol 72:8150-7.

112. Moran, J. V., S. E. Holmes, T. P. Naas, R. J. DeBerardinis, J. D. Boeke, and H. H. Kazazian, Jr. 1996. High frequency retrotransposition in cultured mammalian cells. Cell 87:917-27.

113. Muckenfuss, H., M. Hamdorf, U. Held, M. Perkovic, J. Lower, K. Cichutek, E. Flory, G. G. Schumann, and C. Munk. 2006. APOBEC3 proteins inhibit human LINE-1 retrotransposition. J Biol Chem 281:22161-72.

114. Muckenfuss, H., J. K. Kaiser, E. Krebil, M. Battenberg, C. Schwer, K. Cichutek, C. Munk, and E. Flory. 2007. Sp1 and Sp3 regulate basal transcription of the human APOBEC3G gene. Nucleic Acids Res 35:3784-96.

115. Münk, C., T. Beck, J. Zielonka, A. Hotz-Wagenblatt, S. Chareza, M. Battenberg, J. Thielebein, K. Cichutek, I. G. Bravo, S. J. O'Brien, M. Löchelt, and N. Yuhki. 2008. Functions, structure, and read-through alternative splicing of feline APOBEC3 genes. Genome Biol 9:R48.

116. Münk, C., J. Zielonka, H. Constabel, B. P. Kloke, B. Rengstl, M. Battenberg, F. Bonci, M. Pistello, M. Löchelt, and K. Cichutek. 2007. Multiple restrictions of human immunodeficiency virus type 1 in feline cells. J Virol 81:7048-60.

117. Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, and T. Honjo. 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 102:553-63.

118. Muramatsu, M., V. S. Sankaranand, S. Anant, M. Sugai, K. Kinoshita, N. O. Davidson, and T. Honjo. 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 274:18470-6.

119. Nathans, R., H. Cao, N. Sharova, A. Ali, M. Sharkey, R. Stranska, M. Stevenson, and T. M. Rana. 2008. Small-molecule inhibition of HIV-1 Vif. Nat Biotechnol 26:1187-92.

120. Navarro, F., B. Bollman, H. Chen, R. Konig, Q. Yu, K. Chiles, and N. R. Landau. 2005. Complementary function of the two catalytic domains of APOBEC3G. Virology 333:374-86.

121. Newman, E. N., R. K. Holmes, H. M. Craig, K. C. Klein, J. R. Lingappa, M. H. Malim, and A. M. Sheehy. 2005. Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. Curr Biol 15:166-70.

122. Niewiadomska, A. M., C. Tian, L. Tan, T. Wang, P. T. Sarkis, and X. F. Yu. 2007. Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association. J Virol 81:9577-83.

123. Notredame, C., D. G. Higgins, and J. Heringa. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 302:205-17.

124. Oberste, M. S., and M. A. Gonda. 1992. Conservation of amino-acid sequence motifs in lentivirus Vif proteins. Virus Genes 6:95-102.

125. OhAinle, M., J. A. Kerns, M. M. Li, H. S. Malik, and M. Emerman. 2008. Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. Cell Host Microbe 4:249-59.
126. OhAinle, M., J. A. Kerns, H. S. Malik, and M. Emerman. 2006. Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. J Virol 80:3853-62.
127. Okeoma, C. M., N. Lovsin, B. M. Peterlin, and S. R. Ross. 2007. APOBEC3 inhibits mouse mammary tumour virus replication in vivo. Nature 445:927-30.
128. Opi, S., S. Kao, R. Goila-Gaur, M. A. Khan, E. Miyagi, H. Takeuchi, and K. Strebel. 2007. Human immunodeficiency virus type 1 Vif inhibits packaging and antiviral activity of a degradation-resistant APOBEC3G variant. J Virol 81:8236-46.
129. Park, I. W., K. Myrick, and J. Sodroski. 1994. Effects of vif mutations on cell-free infectivity and replication of simian immunodeficiency virus. J Acquir Immune Defic Syndr 7:1228-36.
130. Peng, G., K. J. Lei, W. Jin, T. Greenwell-Wild, and S. M. Wahl. 2006. Induction of APOBEC3 family proteins, a defensive maneuver underlying interferon-induced anti-HIV-1 activity. J Exp Med 203:41-6.
131. Pery, E., K. S. Rajendran, A. J. Brazier, and D. Gabuzda. 2009. Regulation of APOBEC3 proteins by a novel YXXL motif in human immunodeficiency virus type 1 Vif and simian immunodeficiency virus SIVagm Vif. J Virol 83:2374-81.
132. Petersen-Mahrt, S. K., R. S. Harris, and M. S. Neuberger. 2002. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature 418:99-103.
133. Prochnow, C., R. Bransteitter, M. G. Klein, M. F. Goodman, and X. S. Chen. 2007. The APOBEC-2 crystal structure and functional implications for the deaminase AID. Nature 445:447-51.
134. Purvis, A. 1995. A composite estimate of primate phylogeny. Philos Trans R Soc Lond B Biol Sci 348:405-21.
135. Refsland, E. W., M. D. Stenglein, K. Shindo, J. S. Albin, W. L. Brown, and R. S. Harris. 2010. Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. Nucleic Acids Res.
136. Reich, N., B. Evans, D. Levy, D. Fahey, E. Knight, Jr., and J. E. Darnell, Jr. 1987. Interferon-induced transcription of a gene encoding a 15-kDa protein depends on an upstream enhancer element. Proc Natl Acad Sci U S A 84:6394-8.
137. Retief, J. D. 2000. Phylogenetic analysis using PHYLIP. Methods Mol Biol 132:243-58.
138. Reza Shariflou, M., and C. Moran. 2000. Conservation within artiodactyls of an AATA interrupt in the IGF-I microsatellite for 19-35 million years. Mol Biol Evol 17:665-9.
139. Rogozin, I. B., M. K. Basu, I. K. Jordan, Y. I. Pavlov, and E. V. Koonin. 2005. APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. Cell Cycle 4:1281-5.
140. Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-4.
141. Russell, R. A., and V. K. Pathak. 2007. Identification of Two Distinct HIV-1 Vif Determinants Critical for Interactions with Human APOBEC3G and APOBEC3F. J Virol.
142. Russell, R. A., J. Smith, R. Barr, D. Bhattacharyya, and V. K. Pathak. 2009. Distinct domains within APOBEC3G and APOBEC3F interact with separate regions of human immunodeficiency virus type 1 Vif. J Virol 83:1992-2003.
143. Russell, R. A., H. L. Wiegand, M. D. Moore, A. Schafer, M. O. McClure, and B. R. Cullen. 2005. Foamy virus Bet proteins function as novel inhibitors of the APOBEC3 family of innate antiretroviral defense factors. J Virol 79:8724-31.
144. Santa-Marta, M., F. A. da Silva, A. M. Fonseca, and J. Goncalves. 2005. HIV-1 Vif can directly inhibit apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G-mediated cytidine deamination by using a single amino acid interaction and without protein degradation. J Biol Chem 280:8765-75.
145. Santiago, M. L., M. Montano, R. Benitez, R. J. Messer, W. Yonemoto, B. Chesebro, K. J. Hasenkrug, and W. C. Greene. 2008. Apobec3 encodes Rfv3, a gene influencing neutralizing antibody control of retrovirus infection. Science 321:1343-6.

146. Sawyer, S. L., M. Emerman, and H. S. Malik. 2004. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS Biol 2:E275.
147. Schröfelbauer, B., D. Chen, and N. R. Landau. 2004. A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). Proc Natl Acad Sci U S A 101:3927-32.
148. Schrofelbauer, B., T. Senger, G. Manning, and N. R. Landau. 2006. Mutational alteration of human immunodeficiency virus type 1 Vif allows for functional interaction with nonhuman primate APOBEC3G. J Virol 80:5984-91.
149. Schumacher, A. J., D. V. Nissley, and R. S. Harris. 2005. APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. Proc Natl Acad Sci U S A 102:9854-9.
150. Schwartz, S., W. J. Kent, A. Smit, Z. Zhang, R. Baertsch, R. C. Hardison, D. Haussler, and W. Miller. 2003. Human-mouse alignments with BLASTZ. Genome Res 13:103-7.
151. Shacklett, B. L., and P. A. Luciw. 1994. Analysis of the vif gene of feline immunodeficiency virus. Virology 204:860-7.
152. Shandilya, S. M., M. N. Nalam, E. A. Nalivaika, P. J. Gross, J. C. Valesano, K. Shindo, M. Li, M. Munson, W. E. Royer, E. Harjes, T. Kono, H. Matsuo, R. S. Harris, M. Somasundaran, and C. A. Schiffer. 2010. Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. Structure 18:28-38.
153. Sheehy, A. M., N. C. Gaddis, J. D. Choi, and M. H. Malim. 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 418:646-50.
154. Sheehy, A. M., N. C. Gaddis, and M. H. Malim. 2003. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. Nat Med 9:1404-7.
155. Simon, V., V. Zennou, D. Murray, Y. Huang, D. D. Ho, and P. D. Bieniasz. 2005. Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. PLoS Pathog 1:e6.
156. Smith, J. L., W. Bu, R. C. Burdick, and V. K. Pathak. 2009. Multiple ways of targeting APOBEC3-virion infectivity factor interactions for anti-HIV-1 drug development. Trends Pharmacol Sci 30:638-46.
157. Sova, P., and D. J. Volsky. 1993. Efficiency of viral DNA synthesis during infection of permissive and nonpermissive cells with vif-negative human immunodeficiency virus type 1. J Virol 67:6322-6.
158. Stenglein, M. D., and R. S. Harris. 2006. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. J Biol Chem 281:16837-41.
159. Stenglein, M. D., H. Matsuo, and R. S. Harris. 2008. Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization. J Virol 82:9591-9.
160. Stenglein, S. A., M. S. Rodriguero, E. Chandler, P. Jennings, G. L. Salerno, and P. Nicholson. 2009. Phylogenetic relationships of Fusarium poae based on EF-1alpha and mtSSU sequences. Mycol Res.
161. Stopak, K., C. de Noronha, W. Yonemoto, and W. C. Greene. 2003. HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. Mol Cell 12:591-601.
162. Strebel, K. 2003. Virus-host interactions: role of HIV proteins Vif, Tat, and Rev. Aids 17 Suppl 4:S25-34.
163. Suspene, R., D. Guetard, M. Henry, P. Sommer, S. Wain-Hobson, and J. P. Vartanian. 2005. Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. Proc Natl Acad Sci U S A 102:8321-6.
164. Suyama, M., D. Torrents, and P. Bork. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609-12.
165. Takeda, E., S. Tsuji-Kawahara, M. Sakamoto, M. A. Langlois, M. S. Neuberger, C. Rada, and M. Miyazawa. 2008. Mouse APOBEC3 restricts friend leukemia virus infection and pathogenesis in vivo. J Virol 82:10998-1008.

166. Tan, L., P. T. Sarkis, T. Wang, C. Tian, and X. F. Yu. 2009. Sole copy of Z2-type human cytidine deaminase APOBEC3H has inhibitory activity against retrotransposons and HIV-1. Faseb J 23:279-87.
167. Tanaka, Y., H. Marusawa, H. Seno, Y. Matsumoto, Y. Ueda, Y. Kodama, Y. Endo, J. Yamauchi, T. Matsumoto, A. Takaori-Kondo, I. Ikai, and T. Chiba. 2006. Anti-viral protein APOBEC3G is induced by interferon-alpha stimulation in human hepatocytes. Biochem Biophys Res Commun 341:314-9.
168. Teng, B., C. F. Burant, and N. O. Davidson. 1993. Molecular cloning of an apolipoprotein B messenger RNA editing protein. Science 260:1816-9.
169. Tian, C., X. Yu, W. Zhang, T. Wang, R. Xu, and X. F. Yu. 2006. Differential requirement for conserved tryptophans in human immunodeficiency virus type 1 Vif for the selective suppression of APOBEC3G and APOBEC3F. J Virol 80:3112-5.
170. Tomonaga, K., J. Norimine, Y. S. Shin, M. Fukasawa, T. Miyazawa, A. Adachi, T. Toyosaki, Y. Kawaguchi, C. Kai, and T. Mikami. 1992. Identification of a feline immunodeficiency virus gene which is essential for cell-free virus infectivity. J Virol 66:6181-5.
171. Turelli, P., B. Mangeat, S. Jost, S. Vianin, and D. Trono. 2004. Inhibition of hepatitis B virus replication by APOBEC3G. Science 303:1829.
172. Turelli, P., S. Vianin, and D. Trono. 2004. The innate antiretroviral factor APOBEC3G does not affect human LINE-1 retrotransposition in a cell culture assay. J Biol Chem 279:43371-3.
173. van Regenmortel, M. H., M. A. Mayo, C. M. Fauquet, and J. Maniloff. 2000. Virus nomenclature: consensus versus chaos. Arch Virol 145:2227-32.
174. Virgen, C. A., and T. Hatziioannou. 2007. Antiretroviral activity and Vif sensitivity of rhesus macaque APOBEC3 proteins. J Virol 81:13932-7.
175. Warren, W. C., L. W. Hillier, J. A. Marshall Graves, E. Birney, C. P. Ponting, F. Grutzner, K. Belov, W. Miller, L. Clarke, A. T. Chinwalla, S. P. Yang, A. Heger, D. P. Locke, P. Miethke, P. D. Waters, F. Veyrunes, L. Fulton, B. Fulton, T. Graves, J. Wallis, X. S. Puente, C. Lopez-Otin, G. R. Ordonez, E. E. Eichler, L. Chen, Z. Cheng, J. E. Deakin, A. Alsop, K. Thompson, P. Kirby, A. T. Papenfuss, M. J. Wakefield, T. Olender, D. Lancet, G. A. Huttley, A. F. Smit, A. Pask, P. Temple-Smith, M. A. Batzer, J. A. Walker, M. K. Konkel, R. S. Harris, C. M. Whittington, E. S. Wong, N. J. Gemmell, E. Buschiazzo, I. M. Vargas Jentzsch, A. Merkel, J. Schmitz, A. Zemann, G. Churakov, J. O. Kriegs, J. Brosius, E. P. Murchison, R. Sachidanandam, C. Smith, G. J. Hannon, E. Tsend-Ayush, D. McMillan, R. Attenborough, W. Rens, M. Ferguson-Smith, C. M. Lefevre, J. A. Sharp, K. R. Nicholas, D. A. Ray, M. Kube, R. Reinhardt, T. H. Pringle, J. Taylor, R. C. Jones, B. Nixon, J. L. Dacheux, H. Niwa, Y. Sekita, X. Huang, A. Stark, P. Kheradpour, M. Kellis, P. Flicek, Y. Chen, C. Webber, R. Hardison, J. Nelson, K. Hallsworth-Pepin, K. Delehaunty, C. Markovic, P. Minx, Y. Feng, C. Kremitzki, M. Mitreva, J. Glasscock, T. Wylie, P. Wohldmann, P. Thiru, M. N. Nhan, C. S. Pohl, S. M. Smith, S. Hou, M. Nefedov, et al. 2008. Genome analysis of the platypus reveals unique signatures of evolution. Nature 453:175-83.
176. Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. R. Brent, D. G. Brown, S. D. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. T. Chinwalla, D. M. Church, M. Clamp, C. Clee, F. S. Collins, L. L. Cook, R. R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. T. Dermitzakis, C. Dewey, N. J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D. M. Dunn, S. R. Eddy, L. Elnitski, R. D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G. A. Fewell, P. Flicek, K. Foley, W. N. Frankel, L. A. Fulton, R. S. Fulton, T. S. Furey, D. Gage, R. A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. A. Graves, E. D. Green, S. Gregory, R. Guigo, M. Guyer, R. C. Hardison, D. Haussler, Y. Hayashizaki, L. W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S. Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520-62.

177. Wedekind, J. E., G. S. Dance, M. P. Sowden, and H. C. Smith. 2003. Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. Trends Genet 19:207-16.
178. Wiegand, H. L., and B. R. Cullen. 2007. Inhibition of alpharetrovirus replication by a range of human APOBEC3 proteins. J Virol 81:13694-9.
179. Wiegand, H. L., B. P. Doehle, H. P. Bogerd, and B. R. Cullen. 2004. A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. Embo J 23:2451-8.
180. Wolfe, L. S., B. J. Stanley, C. Liu, W. K. Eliason, and Y. Xiong. 2010. Dissection of the HIV Vif interaction with human E3 ubiquitin ligase. J Virol 84:7135-9.
181. Xiao, Z., E. Ehrlich, Y. Yu, K. Luo, T. Wang, C. Tian, and X. F. Yu. 2006. Assembly of HIV-1 Vif-Cul5 E3 ubiquitin ligase through a novel zinc-binding domain-stabilized hydrophobic interface in Vif. Virology 349:290-9.
182. Xu, H., E. S. Svarovskaia, R. Barr, Y. Zhang, M. A. Khan, K. Strebel, and V. K. Pathak. 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 U S A 101:5652-7.
183. Yamanaka, S., M. E. Balestra, L. D. Ferrell, J. Fan, K. S. Arnold, S. Taylor, J. M. Taylor, and T. L. Innerarity. 1995. Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. Proc Natl Acad Sci U S A 92:8483-7.
184. Yamashita, T., K. Kamada, K. Hatcho, A. Adachi, and M. Nomaguchi. 2008. Identification of amino acid residues in HIV-1 Vif critical for binding and exclusion of APOBEC3G/F. Microbes Infect 10:1142-9.
185. Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555-6.
186. Yu, Q., R. Konig, S. Pillai, K. Chiles, M. Kearney, S. Palmer, D. Richman, J. M. Coffin, and N. R. Landau. 2004. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat Struct Mol Biol 11:435-42.
187. Yu, X., Y. Yu, B. Liu, K. Luo, W. Kong, P. Mao, and X. F. Yu. 2003. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. Science 302:1056-60.
188. Yu, Y., Z. Xiao, E. S. Ehrlich, X. Yu, and X. F. Yu. 2004. Selective assembly of HIV-1 Vif-Cul5-ElonginB-ElonginC E3 ubiquitin ligase complex through a novel SOCS box and upstream cysteines. Genes Dev 18:2867-72.
189. Zennou, V., and P. D. Bieniasz. 2006. Comparative analysis of the antiretroviral activity of APOBEC3G and APOBEC3F from primates. Virology 349:31-40.
190. 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 424:94-8.
191. Zhang, J., and D. M. Webb. 2004. Rapid evolution of primate antiviral enzyme APOBEC3G. Hum Mol Genet 13:1785-91.
192. Zhang, L., B. Ma, L. Wang, and Y. Xu. 2003. Greedy method for inferring tandem duplication history. Bioinformatics 19:1497-504.
193. Zhang, W., G. Chen, A. M. Niewiadomska, R. Xu, and X. F. Yu. 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:e3963.
194. Zheng, Y. H., D. Irwin, T. Kurosu, K. Tokunaga, T. Sata, and B. M. Peterlin. 2004. Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. J Virol 78:6073-6.
195. Zielonka, J., I. G. Bravo, D. Marino, E. Conrad, M. Perković, M. Battenberg, K. Cichutek, and C. Münk. 2009. Restriction of equine infectious anemia virus by equine APOBEC3 cytidine deaminases. J Virol 83:7547-59.

APPENDIX I

Guidelines for Naming Non-Primate APOBEC3 Genes and Proteins

Reprinted with permission from: R. S. LaRue, Andrésdóttir V, Blanchard Y, Conticello SG, Derse D, Emerman M, Greene WC, Jónsson SR, Landau NR, Löchelt M, Malik HS, Malim MH, Münk C, O'Brien SJ, Pathak VK, Strebel K, Wain-Hobson S, Yu XF, Yuhki N, Harris RS. (2009). J Virol. 2009 Jan;83(2):494-7.

Authors' contributions

R. S. LaRue and R.S. Harris composed the new A3 nomenclature and wrote the manuscript. All remaining authors supported new A3 nomenclature and contributed to the editing of the manuscript.

APOBEC3 genes are unique to mammals but copy numbers vary significantly

APOBEC3 (A3) proteins are of considerable interest because most are potent DNA cytosine deaminases that have the capacity to restrict the replication and/or edit the sequences of a wide variety of parasitic elements, including many retroviruses and retrotransposons [reviewed by (4, 7-9, 12)]. Likely substrates include i) lentiviruses such as HIV-1, HIV-2, SIV, MVV, FIV and EIAV, ii) alpha-, beta-, gamma- and delta-retroviruses such as RSV, MPMV/MMTV, MuLV/FeLV and HILV/BLV, respectively, iii) spumaviruses such as PFV and FFV, iv) hepadnaviruses such as HBV, v) endogenous retroviruses and LTR-retrotransposons such as HERV-K, IAP, MusD and PERV, vi) non-LTR retrotransposons such as L1 and Alu and vii) DNA viruses such as AAV and HPV. Over the past few years, there has also been an increasing appreciation for the multiple, distinct mechanisms that parasitic elements use to co-exist with the A3 proteins of their hosts. Together, this indicates that the evolution of the A3 proteins has been driven by a requirement to minimize the spread of exogenous and endogenous genetic threats. The likelihood that the A3 proteins might exist solely for this purpose has been supported recently by studies indicating that A3-deficient mice have no obvious phenotypes, apart from a notable increase in susceptibility to retrovirus infection (13, 16, 18, 20).

The *A3* genes are specific to mammals and are organized in a tandem array between two vertebrate-conserved flanking genes *CBX6* and *CBX7* [**Fig 1A;** e.g., (11)]. Based on a limited number of genomic sequences, it is already clear that *A3* copy number can vary greatly from mammal to mammal. For instance, mice have one *A3* gene (9, 13), pigs have

two (11), cattle and sheep have three (11), cats have four (14), horses have six (2) and humans and chimpanzees have seven (5, 9, 10). Other mammals are likely to have copy numbers within this range but the cat and horse loci, in particular, highlight the difficulty in making such predictions (2, 14).

Each *APOBEC3* gene is comprised of one or two zinc-coordinating domains

Naming the mammalian *A3* genes is complicated further by fact that that each gene encodes a single- or a double zinc(Z)-coordinating domain protein. For instance, human *A3A*, *A3C* and *A3H* encode single Z-domain proteins, whereas human *A3B*, *A3DE*, *A3F* and *A3G* encode double Z-domain proteins. The Z domain is required for catalytic activity, but some domains have not elicited activity and can therefore be regarded as pseudo-catalytic. Nevertheless, all Z domains can be readily identified by four invariant residues, one histidine, one glutamate and two cysteines, organized H-x_1-E-x_{23-28}-C-x_{2-4}-C (x can be nearly any one of twenty amino acids; see **Fig 1B** and below). The histidine and two cysteines are required to bind a single zinc atom, and, at least for the catalytic domains, the glutamate is predicted to promote the formation of a hydroxide ion required for deamination.

Each Z domain clearly belongs to one of three distinct phylogenetic clusters – originally termed Z1b, Z1a and Z2 [(6); adopted by (6, 15, 17)]. However, while we acknowledge the logical nature of these Z-based groupings, we propose a simplification of the scheme

to Z1, Z2 and Z3, respectively. This minor nomenclature change was motivated because i) lower case letters are needed to help describe unique A3 variants (see below), ii) a key mammalian ancestor likely had a *CBX6-Z1-Z2-Z3-CBX7* locus organization (11), and iii) the Z3 domain has so far been found to be invariably located at the distal end of the locus next to *CBX7* (**Fig 1A**).

Z domain assignments can be made simply by scanning predicted polypeptide sequences for key identifying residues (**Fig 1B**). This 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 (3). Z2 domains possess a unique tryptophan-phenylalanine (WF) motif five residues after the (pseudo)catalytic glutamate. Finally, Z3 domains have a TW-S-C-x_{2-4}-C zinc-coordinating motif, whereas both the Z1 and Z2 domains have a SW-S-C-x_{2-4}-C motif. Since many A3 proteins have been subject to positive selection (19), 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.

However, although these simple rules enable initial Z domain assignments, it should be noted that several other differences combine to distinguish each of the three Z types and final assignments should be verified by comprehensive phylogenetic analyses. One should also be aware of the fact that the mammalian *A3* locus is frequently involved in genetic recombination events such as unequal cross-overs (leading to deletions or insertions) and gene conversions [*e.g.*, (11)]. Thus, to minimize the potentially confounding effects of recombination, we further recommend (at least for the purposes of

nomenclature) that *A3* gene descriptions be based exclusively on the Z domain assignments [*i.e.*, based on phlyogenetic analyses of the Z domain-encoding exon; *e.g.*, **Fig 1A** and (11)].

A Z-domain-based nomenclature system for non-primate *APOBEC3* genes

With new technologies delivering tidal waves of genomic and transcribed sequences to the scientific community, it is important to have nomenclature systems in place to facilitate the annotation, dissemination and comparison of specific genes and gene families. The current Human Genome Organization (HUGO) conventions suggest that the human gene name be used to annotate the orthologous genes of non-human species (www.genenames.org). The HUGO system can be applied readily to the *A3* genes of primates such as the chimpanzee and the rhesus macaque, which nearly aligns domain-for-domain with the human *A3* locus (**Fig 1a**). However, the *A3* loci of non-primate mammals pose a particularly difficult problem because they vary in size, Z domain type and Z domain organization. Read-through transcription, alternative splicing and internal transcriptions initiation further complicate naming schemes [e.g., (11, 14)]. Most importantly, it is impossible (and incorrect) to deduce orthologous relationships between humans and non-primate mammals, because each species' A3 proteins are the product of a unique, divergent evolutionary history that was shaped by immeasurable selective pressures.

Therefore, to simplify matters, we propose the following Z domain-based nomenclature system that can be applied easily to annotate and describe the *APOBEC3* repertoire of any non-primate mammal. It is based on the fact that the *A3* genes are clearly modular in nature, consisting of one (Z1, Z2 or Z3) or some combination of two Z domains [Z2-Z1, Z2-Z2, Z2-Z3; *e.g.*, (2, 11, 14)] Other combinations may very well exist but they have yet to be described, This Z domain-based system is best applied once a species' entire *A3* genomic locus has been determined, and it does not immediately require knowledge of mRNA or protein coding capacity.

First, once an *A3* locus has been sequenced (ideally, completely), the Z domain type should be assigned as described above. A simple example is the cow *A3* locus, which consists of three distinct Z domains organized Z1-Z2-Z3 (11). A more complex example is the horse, consisting of two Z1 domains, five Z2 domains and a single Z3 domain (2). Second, in such an instance when multiple domains of a single Z type exist, we propose that lower case letters be used to distinguish each distinct domain (ideally applied starting at the *CBX6* side of the locus and ending at the *CBX7* side; *i.e.*, starting at the 5' end). For instance, eight Z domain horse *A3* repertoire would be designated Z1a-Z1b-Z2a-Z2b-Z2c-Z2d-Z2e-Z3. Finally, based on mRNA expression data, which will undoubtedly reveal how the Z domains mix and match *in vivo*, additional assignments can be made. Single Z domain genes, mRNAs and proteins can be annotated simply by adding the APOBEC3 (A3) prefix. For instance the cow has A3Z1, A3Z2 and A3Z3. Double Z domain genes, mRNAs and proteins can be annotated by adding an A3 prefix and merging the Z domain designations. For instance, cattle have three *APOBEC3* genes:

A3Z1, *A3Z2* and *A3Z3* (11). Following the logic, double Z domain genes, mRNAs and proteins can be annotated by adding the A3 prefix and pairing the Z domain designations. For instance, the cow also has an A3Z2-Z3 protein (11), and the coding potential of the horse *A3* repertoire can be described as A3Z1a, A3Z1b, A3Z2a-Z2b, A3Z2c-Z2d, A3Z2e and A3Z3 [(2) and see **Fig 1A**]. New names are listed in **Table 1** for all of the *A3* genes of non-primate mammals whose *A3* genomic loci are 'complete'.

At first glance, this new nomenclature system may appear cumbersome. However, we suspect that continual exposure and practice will yield both familiarity and, possible, a colloquial 'short form' lacking common denominators. Again, using the cow and horse as examples, the former has Z1, 2, 3, and 2-3 types of A3 proteins and the latter has Z1a, 1b, 2ab, 2cd, 2e and 3 types of A3 proteins.

It also is worth mentioning that a Z domain-based system is also possible for the primate A3s (e.g., **Fig 1A**). A complete conversion to this system would certainly facilitate intra Z-type and inter-species comparisons, but we fully recognize that the well-established (and popular_ human APOBEC3A through –H designations are not likely to be superseded (**Fig 1A**). We further recognize that the mouse may also be a special case, because the generic *A3* designation has already been used to describe its single (albeit double Z domain) gene. However, regardless of whether the new nomenclature scheme is adopted, it is important to emphasize again that it guards against the false implication of orthology between certain human *A3* genes and the *A3* genes found in other mammals. Previously, *A3* genes have been tentatively named on the basis of BLAST-score matches,

which have been shown to be a notoriously poor means of establishing orthology, especially when reciprocal best-BLAST hits are not employed. This, the new nomenclature scheme is not only simple and logical, it is also more formally correct than current schemes.

Finally, it is important to point out that the new system readily accommodates *A3* variants created by read-through transcription and alternative splicing. For instance, the feline *A3* locus can now be designated *A3Z2a-A3Z2b-A3Z2c-A3Z3*, which encodes four similarly designated single domain proteins and a novel A3Z2b-Z3 variant (14). Moreover, a numeric suffix can be added to each designation to accommodate splice variants. Overall, we hope that the intrinsic logic of the simplified Z domain-based nomenclature system will enable the mammalian *A3* genes to be fully described and appropriately included in a wealth of comparative studies to better understand a broad range of host-pathogen conflicts.

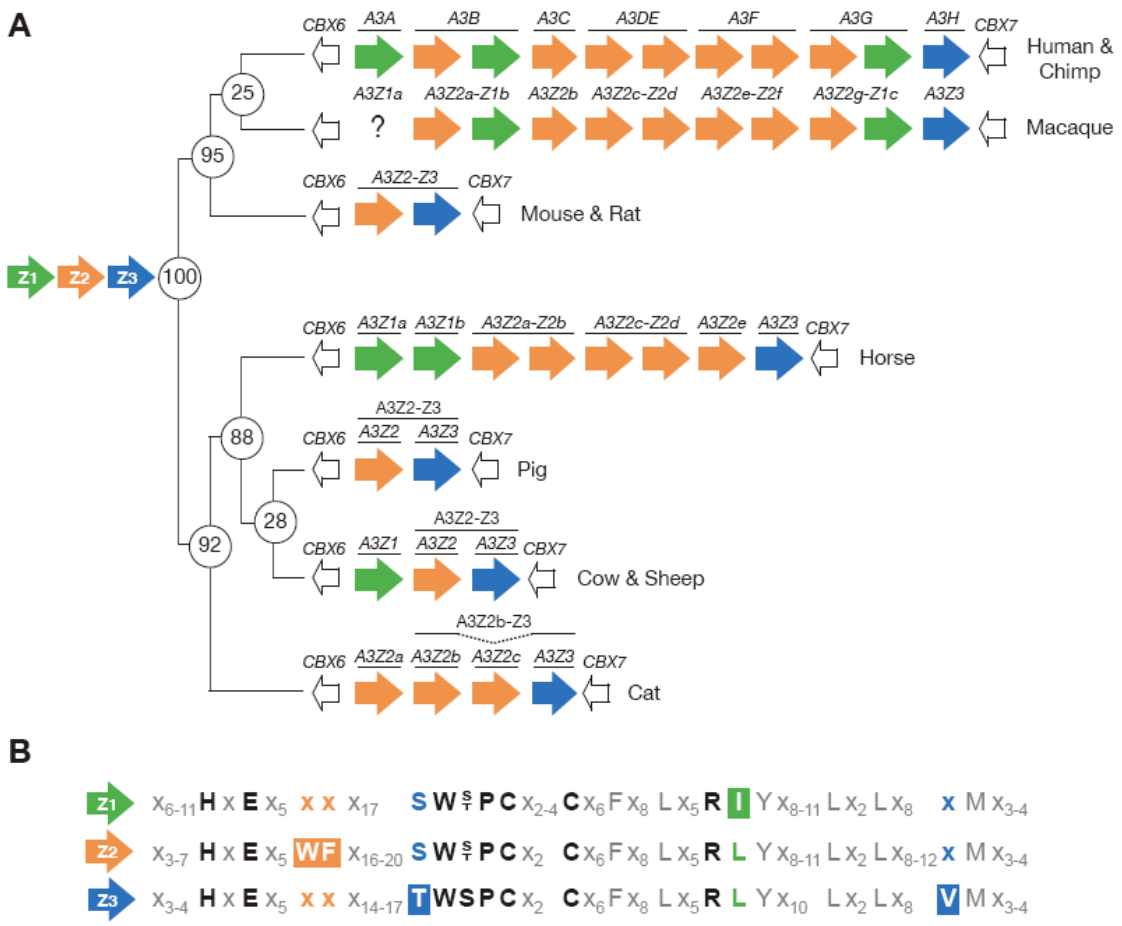


FIG. 1. (**A**) Schematics of the *A3* repertoires of mammals whose genomes have been sequenced. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. For all of the indicated species (and likely all mammals), *CBX6* is located immediately upstream and *CBX7* downstream of the *A3* locus. Either macaque *A3A* does not exist or its genomic sequence is not quite complete. The inferred ancestral *A3* repertoire was deduced through comparative studies (11). The numbers at the phylogenetic tree 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 (1). (**B**) Highlights of amino acid conservation among the three distinct Z domain groups and within each individual group [based on multiple sequence alignments (11)]. Residues discussed in the text are colored or emboldened, and other notable residues are shaded grey. An 'X' specifies nearly any amino acid.

Table 1. The APOBEC3 genes and proteins of representative non-primate mammals.

Genus species (common name)	**Old names** (reference)		**New names** (reference)	
	Gene [a]	**Protein** [a]	**Gene** [b]	**Protein**
Bos taurus (cow)			*A3Z1* (12)	A3Z1
			A3Z2 (12)	A3Z2
			A3Z3 (12)	A3Z3
	A3F (11)	A3F		A3Z2-Z3 (12)
Equus caballus (horse)	*A3A1* (2)	A3A1	*A3Z1a*	A3Z1a
	A3A2 (2)	A3A2	*A3Z1b*	A3Z1b
	A3F1 (2)	A3F1	*A3Z2a-Z2b*	A3Z2a-Z2b
	A3F2 (2)	A3F2	*A3Z2c-Z2d*	A3Z2c-Z2d
	A3C (2)	A3C	*A3Z2e*	A3Z2e
	A3H (2)	A3H	*A3Z3*	A3Z3
Felis catus (cat)	*A3Cc* (16)	A3Cc	*A3Z2a*	A3Z2a
	A3Ca (16)	A3Ca	*A3Z2b*	A3Z2b
	A3Cb (16)	A3Cb	*A3Z2c*	A3Z2c
	A3H (16)	A3H	*A3Z3*	A3Z3
		A3CH (16)		A3Z2b-Z3
Mus musculus (mouse)	*A3* (14)	A3	*A3Z2-Z3*	A3Z2-Z3
Ovis aries (sheep)			*A3Z1* (12)	A3Z1
			A3Z2 (12)	A3Z2
			A3Z3 (12)	A3Z3
	A3F (11)	A3F		A3Z2-Z3 (12)
Rattus norvegicus (rat)	*A3*	A3	*A3Z2-Z3*	A3Z2-Z3
Sus scrofa (pig)			*A3Z2* (12)	A3Z2
			A3Z3 (12)	A3Z3
	A3F (11)	A3F		A3Z2-Z3 (12)

[a] Some cells have been left empty because the new gene and protein names proposed here will also be used in corresponding original research articles (12).

[b] The cells for some of the gene names have been left empty because an argument can be made that the resulting double Z domain protein is the product of two distinct genes, created by read-through transcription and alternative splicing [*e.g.*, (12, 16)].

References

1. **Bininda-Emonds, O. R., M. Cardillo, K. E. Jones, R. D. MacPhee, R. M. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis.** 2007. The delayed rise of present-day mammals. Nature **446:**507-12.
2. **Bogerd, H. P., R. L. Tallmadge, J. L. Oaks, S. Carpenter, and B. R. Cullen.** 2008. Equine infectious anemia virus resists the antiretroviral activity of equine APOBEC3 proteins through a packaging-independent mechanism. J Virol **82:**11889-901.
3. **Chen, K. M., E. Harjes, P. J. Gross, A. Fahmy, Y. Lu, K. Shindo, R. S. Harris, and H. Matsuo.** 2008. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. Nature **452:**116-9.
4. **Chiu, Y. L., and W. C. Greene.** 2008. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annu Rev Immunol **26:**317-53.
5. **Consortium, T. C. S. a. A.** 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature **437:**69-87.
6. **Conticello, S. G., C. J. Thomas, S. K. Petersen-Mahrt, and M. S. Neuberger.** 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol **22:**367-77.
7. **Cullen, B. R.** 2006. Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors. J Virol **80:**1067-76.
8. **Goila-Gaur, R., and K. Strebel.** 2008. HIV-1 Vif, APOBEC, and intrinsic immunity. Retrovirology **5:**51.
9. **Harris, R. S., and M. T. Liddament.** 2004. Retroviral restriction by APOBEC proteins. Nat Rev Immunol **4:**868-77.
10. **Jarmuz, A., A. Chester, J. Bayliss, J. Gisbourne, I. Dunham, J. Scott, and N. Navaratnam.** 2002. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. Genomics **79:**285-96.
11. **LaRue, R. S., S. R. Jónsson, K. A. Silverstein, M. Lajoie, D. Bertrand, N. El-Mabrouk, I. Hötzel, V. Andrésdóttir, T. P. Smith, and R. 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 Mol Biol **9:**104.
12. **Malim, M. H., and M. Emerman.** 2008. HIV-1 accessory proteins--ensuring viral survival in a hostile environment. Cell Host Microbe **3:**388-98.
13. **Mikl, M. C., I. N. Watt, M. Lu, W. Reik, S. L. Davies, M. S. Neuberger, and C. Rada.** 2005. Mice deficient in APOBEC2 and APOBEC3. Mol Cell Biol **25:**7270-7.
14. **Münk, C., T. Beck, J. Zielonka, A. Hotz-Wagenblatt, S. Chareza, M. Battenberg, J. Thielebein, K. Cichutek, I. G. Bravo, S. J. O'Brien, M. Löchelt, and N. Yuhki.** 2008. Functions, structure, and read-through alternative splicing of feline APOBEC3 genes. Genome Biol **9:**R48.
15. **OhAinle, M., J. A. Kerns, H. S. Malik, and M. Emerman.** 2006. Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. J Virol **80:**3853-62.
16. **Okeoma, C. M., N. Lovsin, B. M. Peterlin, and S. R. Ross.** 2007. APOBEC3 inhibits mouse mammary tumour virus replication in vivo. Nature **445:**927-30.
17. **Rogozin, I. B., M. K. Basu, I. K. Jordan, Y. I. Pavlov, and E. V. Koonin.** 2005. APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. Cell Cycle **4:**1281-5.
18. **Santiago, M. L., M. Montano, R. Benitez, R. J. Messer, W. Yonemoto, B. Chesebro, K. J. Hasenkrug, and W. C. Greene.** 2008. Apobec3 encodes Rfv3, a gene influencing neutralizing antibody control of retrovirus infection. Science **321:**1343-6.
19. **Sawyer, S. L., M. Emerman, and H. S. Malik.** 2004. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS Biol **2:**E275.

20. **Takeda, E., S. Tsuji-Kawahara, M. Sakamoto, M. A. Langlois, M. S. Neuberger, C. Rada, and M. Miyazawa.** 2008. Mouse APOBEC3 restricts friend leukemia virus infection and pathogenesis in vivo. J Virol **82:**10998-1008.

APPENDIX II

Links to permissions and license agreements for published articles

Chapter 2

The Artiodactyl *APOBEC3* Innate Immune Repertoire Shows Evidence for a Multi-functional Domain Organization that Existed in the Ancestor of Placental Mammals

Reprinted with permission from: R.S. LaRue, S.R. Jónsson, K.A.T. Silverstein, M. Lajoie, D. Bertrand, N. El-Mabrouk, I. Hötzel, V. Andrésdóttir, T.P.L. Smith, and R.S. Harris (2008) BMC Mol. Biol.

BioMed Central copyright and license agreement

http://www.biomedcentral.com/info/authors/license

Chapter 3

Lentiviral Vif Degrades the APOBEC3Z3/APOBEC3H Protein of its Mammalian Host and is Capable of Cross-Species Activity

Reprinted with permission from: R. S. LaRue, J. Lengyel, S.R. Jónsson, V. Andrésdóttir, and R. S. Harris. (2010) J. of Virol.

AMERICAN SOCIETY FOR MICROBIOLOGY LICENSE TERMS AND CONDITIONS

https://s100.copyright.com/CustomerAdmin/PLF.jsp?lID=2010071_1280515427541

Appendix I

Guidelines for Naming Non-Primate APOBEC3 Genes and Proteins

Reprinted with permission from: R. S. LaRue, Andrésdóttir V, Blanchard Y, Conticello SG, Derse D, Emerman M, Greene WC, Jónsson SR, Landau NR, Löchelt M, Malik HS, Malim MH, Münk C, O'Brien SJ, Pathak VK, Strebel K, Wain-Hobson S, Yu XF, Yuhki N, Harris RS. (2009). J Virol. 2009 Jan;83(2):494-7.

AMERICAN SOCIETY FOR MICROBIOLOGY LICENSE TERMS AND CONDITIONS

https://s100.copyright.com/CustomerAdmin/PLF.jsp?lID=2010071_1280515702269