The ancient and recurrent evolution of Apobec3 antiviral activity in primates

Nisha K. Duggal

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2012

Reading Committee:
Michael Emerman, Chair
Adam Geballe
Joshua Akey

Program Authorized to Offer Degree:
Molecular and Cellular Biology

University of Washington

Abstract

The ancient and recurrent evolution of Apobec3 antiviral activity in primates

Nisha K. Duggal

Chair of the Supervisory Committee:
Affiliate Professor Michael Emerman
Microbiology

Host restriction factors are potent, widely expressed, intracellular blocks to viral replication that are an important component of the innate immune response to viral infection. However, viruses have evolved mechanisms of antagonizing restriction factors. Through evolutionary pressure for both host survival and virus replication, an evolutionary "arms race" has developed that drives continuous rounds of selection for beneficial mutations in restriction factor genes and their viral antagonists. Because viruses can evolve faster than their hosts, the modern-day vertebrate innate immune system is optimized, for the most part, to defend against ancient viruses, rather than current viral threats. Thus, the evolutionary history of restriction factors might, in part, explain why humans are susceptible or resistant to the viruses present in the modern world.

In primates, the Apobec3 family of host restriction factors is a group of seven cytidine deaminases that inhibit the replication of retroviruses and retrotransposons. Here, I characterize the antiviral activity of Apobec3DE across primate species and of Apobec3 family members across human populations. First, I show that human and chimpanzee orthologues of Apobec3DE inhibit retrotransposons that are currently active in hominids. Chimpanzee Apobec3DE also potently restricts lentiviruses, unlike human Apobec3DE, which has weaker antiviral activity against retroviruses. I also show that Apobec3DE evolved rapidly in chimpanzee ancestors, and

this ancient evolution drove the increased antiviral activity of chimpanzee Apobec3DE against modern-day viruses.

Next, I evaluate novel variants of Apobec3 genes in diverse human populations. I show that polymorphisms in Apobec3DE, Apobec3F, and Apobec3G decrease endogenous expression or antiviral activity against HIV-1. Interestingly, I find that Apobec3DE has evolved under purifying selection in humans, suggesting that mutations in Apobec3DE may be weakly deleterious. However, because the current level of Apobec3DE antiviral activity against HIV-1 is not optimal, Apobec3DE likely evolved to recognize an ancient virus unlike HIV-1. Together, this data suggests that Apobec3DE is likely to play a role in host defense against retroelements in primates and may alter our susceptibility to viruses.

Table of Contents

List of Figures

List of Tables

Chapter I

Evolutionary conflicts between viruses and restriction factors shape immunity

Introduction

Restriction factors are proteins of the innate immune system encoded in the germline genome that inhibit the replication of viruses during their lifecycle in host cells. These host proteins are dedicated antiviral factors that are often induced by interferon signaling as part of the innate immune response, are antagonized by viral factors, and are rapidly evolving. The term "restriction factor" was historically adopted by labs studying retroviruses as a result of the discovery of the *Fv1* locus in mice that conferred resistance to murine retroviruses in the 1960s (Lilly, 1967; Pincus et al., 1971). However, this term can also be applied more broadly to host-encoded gene products that inhibit the intracellular replication of any animal virus. Recent work has shown that host susceptibility to viral infection and disease is determined by a summation of the components of the innate immune system (such as restriction factors) and the viral proteins that have evolved to evade or destroy these host defenses. In this chapter, I will describe the general characteristics of restriction factors and show how the evolutionary conflict between viruses and restriction factors has shaped our modern immune systems. These topics are of particular relevance today because of many new discoveries of restriction factors and determinants of viral susceptibility that have rapidly gained attention. I will also describe the Apobec3 family of restriction factors in more detail, as they are the focus of my research.

Characteristics of restriction factors

Classical innate immunity against viruses is mediated by specialized cells such as natural killer (NK) cells, dendritic cells, and macrophages. By contrast, restriction factors are germline-encoded factors that mediate a “cell-intrinsic" immune response. They are part of the broader innate immune repertoire of cellular molecules that detect and respond to viral infections in the absence of previous exposure. Typically, viral infections are detected by cytoplasmic pathogen recognition receptors (PRRs) or Toll-like receptors (TLRs), which trigger an interferon (IFN) response that induces a program of interferon-stimulated genes (ISGs) with broad-ranging effects on cell growth and metabolism (reviewed in (Takaoka and Yanai, 2006; Wilkins and Gale, 2010)). Many of these ISGs are restriction factors that specifically inhibit viral growth within the infected cell. Table 1 lists the general features of the restriction factors that target retroviruses and other viruses that are described in this chapter. Table 1 is not a comprehensive list of restriction factors but contains some of the best-studied examples.

There are several distinguishing characteristics of restriction factors that allow one to make inferences about their role in the evolution of both the host and the virus. Typically, a host gene can be defined as a restriction factor if it encodes a protein that: has antiviral activity as its major biological function; is induced by interferon or by virus infection; is antagonized by a viral protein; and shows evolutionary signatures of genetic conflict (positive selection). The majority of true restriction factors share these features, as described in detail below. However, the exceptions to these definitions are also highlighted, as they can be enlightening with regard to understanding the additional cellular roles that restriction factors may play.

Restriction factor	Viral targets	IFN induced	Viral lifecycle stage inhibited	Viral antagonists	Under positive selection
Fv1	Retroviruses	No	Capsid uncoating	None known	Yes
TRIM5α and TRIM-CYP	Retroviruses	Yes	Capsid uncoating	None known (escape though Capsid mutations)	Yes
Apobec3A-H	Retroviruses Retrotransposons Hepadnaviruses	Yes	Reverse transcription	Vif (lentiviruses) Bet (spumaviruses) Gag (gammaretroviruses)	Apobec3G, Apobec3H
SAMHD1	Retroviruses	Yes	Reverse transcription	Vpx (some SIVs) Vpr (some SIVs)	Yes
ZAP	Retroviruses Filoviruses Alphaviruses	Yes	Viral protein translation	None known	Yes
Tetherin	Retroviruses Flaviviruses Herpesviruses Rhabdoviruses Paramyxoviruses Arenaviruses	Yes	Budding	Nef (some SIVs) Vpu (HIV-1) Env (HIV-2) Glycoprotein (EBOV) K5 (KSHV)	Yes
Viperin	Orthomyxoviruses Flaviriruses Herpesviruses Alphaviruses Paramyxoviruses	Yes	Budding	None known	Yes
MxA and Mx1	Orthomyxoviruses Paramyxoviruses Picornaviruses Hepadnaviruses Alphaviruses Rhabdoviruses Bunyaviruses Togaviruses	Yes	Nucleocapsid transport or other early lifecycle step	None known	ND
IFITM1-3	Orthomyxoviruses Flaviviruses Coronaviruses	Yes	Endosomal fusion/uncoating	None known	ND
PKR	Poxviruses Herpesviruses	Yes	Viral protein translation	TRS1 and IRS1 (HCMV) K3L and E3L (Vaccinia Virus) and others	Yes

Table 1. Characteristics of selected restriction factors.
SIV: simian immunodeficiency virus, EBOV: Ebola virus, KSHV: Kaposi's sarcoma-associated herpesvirus, HCMV: Human cytomegalovirus, ND: not determined

Activity and expression of viral restriction factors

Many restriction factors are IFN-stimulated genes (Table 1), which is consistent with their fundamental role in an antiviral response. The IFN-mediated induction of many restriction factors is also an indication that their major activity is to combat pathogens, rather than some central metabolic or developmental role in the organism. Moreover, as many restriction factors cause destructive events such as protein modifications or nucleotide mutations, their expression needs to be tightly controlled to avoid deleterious effects on cell growth in the absence of viral challenge. However, IFN induction is not a universal property of restriction factors, as some are also expressed constitutively. In cases where protein expression is constitutive, it is probable that the restriction factor has a role in restricting endogenous events as well. For example, in the Apobec3 family of cytidine deaminases, Apobec3G is constitutively expressed by many cell types including T cells and germ cells (Koning et al., 2009; Refsland et al., 2010). Although it has a well-characterized role in T cells of inhibiting retroviruses through hypermutation of viral genomes during reverse transcription, Apobec3G might have an even more ancient role in protecting the host genome in germ cells from endogenous retrotransposons, which do not lead to IFN responses (Stetson, 2009).

During an acute viral infection, each productively infected cell generates many infectious particles, leading to exponential viral growth. Therefore, restriction factors must have extremely potent antiviral activity to have any significant effect on viral loads (restriction factors included in Table 1, for example, decrease viral infectivity by ten-fold or more). This antiviral activity can be demonstrated experimentally by over-expression of a restriction factor, which causes a decrease in virus growth, or knockdown of a restriction factor, which causes an increase in virus growth. For example, SAMHD1 was recently defined as a novel restriction factor that is present

in monocytes because decreasing its endogenous expression in monocytes using RNAi enhanced the replication of HIV-1 and exogenously expressing SAMHD1 cells in epithelial cells restricted HIV-1 replication (Hrecka et al., 2011; Laguette et al., 2011). Additionally, the antiviral activity of restriction factors is specific to families of viruses. For example, TRIM5α seems to be active only against retroviruses because it inhibits viral replication by means of a specific interaction with retroviral capsid proteins (Johnson and Sawyer, 2009). By contrast, tetherin can restrict enveloped viruses across several virus families because it non-specifically incorporates into the cell and virus membranes and prevents efficient viral release by tethering enveloped viruses to the cell (Table 1) (Le Tortorec et al., 2011).

The major biological activity of restriction factors is to inhibit viral replication. In many cases where restriction factor function can be examined by gene knockout in mice, ablation of the restriction factor has no untoward effect on mouse development. For example, knockout mice of the single murine *Apobec3* gene are viable, and the only reported phenotypes is that they are more susceptible to murine retroviruses than are their wild-type counterparts (Okeoma et al., 2007). In fact, natural mutations in *Apobec3* and the *Mx* locus that abolish function exist in some inbred mouse strains (Haller et al., 1987; Okeoma et al., 2009). Similarly, mice with natural or engineered mutations in mouse *Tetherin*, *Viperin*, or *Ifitm3* genes are also viable but are more sensitive to some viral infections (Barrett et al., 2012; Everitt et al., 2012; Liberatore and Bieniasz, 2011; Swiecki et al., 2012; Tan et al., 2012). However, it is possible that some restriction factors have additional cellular roles other than viral restriction. For example, TRIM5α has a more general role in antiviral signaling in addition to its specific role in retroviral restriction (Pertel et al., 2011; Tareen and Emerman, 2011), and mutations in human *SAMHD1* are associated with autoimmune disease (Rice et al., 2009). However, perhaps as a result of the

duplication of many restriction factors within a host (described further below), restriction factors can undergo sub-functionalization, in which one gene retains an essential cellular function whereas its paralogue becomes a dedicated antiviral factor.

Viral antagonists of restriction factors

Viruses have evolved antagonists to restriction factors. These viral proteins are often encoded by 'accessory genes' that are not needed for virus replication except in the presence of restriction factors (Malim and Emerman, 2008). Restriction factors, such as Tetherin, that inhibit the replication of multiple virus families can be antagonized by diverse viral proteins from different viral families (Table 1). In cases where there is no known viral antagonist to a particular restriction factor, it is possible that the virus can escape restriction by mutation of the viral protein targeted by a restriction factor, as is the case for lentiviral evasion of TRIM5α-mediated restriction by viral capsid mutations (Hatziioannou et al., 2006; Kamada et al., 2006). It is also formally possible that a newly evolved restriction factor might not yet have selected for a viral antagonist. However, in most cases, the inability to identify a viral antagonist is more likely attributed to the fact that the relevant sets of viruses and host species have yet to be examined. Viral antagonists can overcome restriction factors using several different mechanisms. For example, viral antagonists can couple the restriction factor to protein degradation pathways, cause the mislocalization of the restriction factor and hence downregulate functional expression, or function as mimics of the restriction factor substrate (Figure 1).

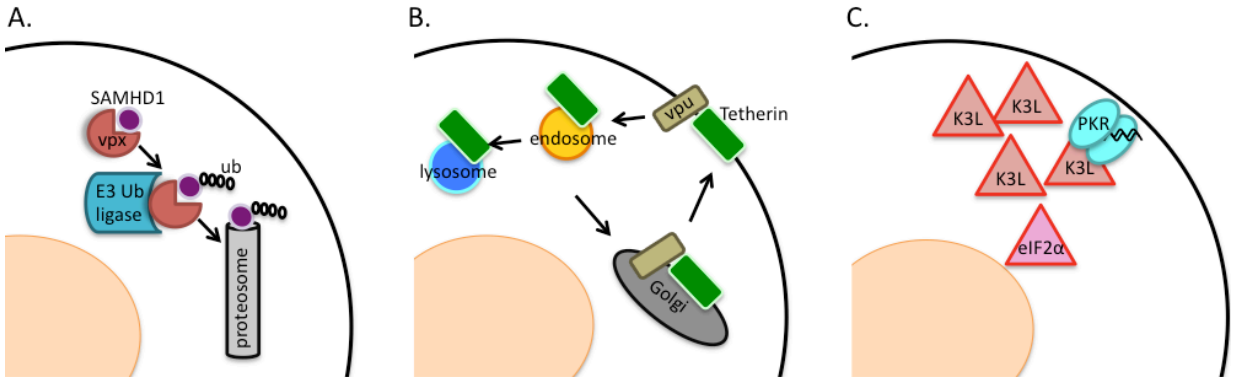


Fig. 1 Mechanisms of viral antagonism of host restriction factors.
(A) Degradation. The lentiviral accessory protein Vpx antagonizes the host restriction factor SAMHD1 by targeting it for degradation (Hrecka et al., 2011; Laguette et al., 2011). Vpx bridges SAMHD1 to an E3 Ubiquitin (Ub) ligase complex, which ubiquitylates SAMHD1, thus targeting SAMHD1 for degradation by the proteasome. (B) Mislocalization. The HIV-1 accessory protein Vpu antagonizes the host restriction factor Tetherin by mislocalization leading to functional downregulation (Kueck and Neil, 2012). Vpu interacts with tetherin at the plasma membrane and trafficks it to the early endosome. Tetherin is then sequestered in the trans-Golgi network (TGN), where it is unable to restrict viral budding from the cell surface, or degraded in the lysosome. (C) Mimicry. The poxvirus accessory protein K3L antagonizes the host PKR pathway by acting as a mimic of the PKR substrate, eIF2α (Elde and Malik, 2009). PKR is activated by binding to double-stranded RNA and induces an antiviral signaling pathway that leads to the inhibition of protein translation. By competing with eIF2α for binding to PKR, K3L prevents eIF2α from inhibiting translation.

To antagonize the restriction factor SAMHD1, the Vpx protein encoded by HIV-2 and related primate lentiviruses targets SAMHD1 for ubiquitylation followed by proteasomal degradation (Fig. 1a) by simultaneously binding SAMHD1 and an adapter protein in the Cul4 ubiquitin ligase complex (Hrecka et al., 2011; Laguette et al., 2011). The lentiviral Vif protein antagonizes Apobec3G by a similar mechanism (Mehle et al., 2004; Yu et al., 2003). By contrast, the lentiviral Vpu protein antagonizes the restriction factor Tetherin by altering its normal cellular localization (Fig. 1b). By a direct protein-protein interaction, Vpu sequesters Tetherin in the trans-Golgi network and re-directs it from the cell membrane to the endosome, where it is unable to restriction viral budding from the cell membrane (Kueck and Neil, 2012). A third mechanism of antagonism is illustrated by K3L, a poxvirus-encoded antagonist of the host antiviral protein kinase R (PKR) pathway. Upon recognition of double-stranded RNA, PKR

inhibits protein translation by phosphorylating eukaryotic initiation factor subunit 2-α (eIF2α). K3L is structurally homologous to eIF2α and competes for binding to PKR (Fig. 1c). By acting as a mimic of eIF2α, K3L prevents the activation of eIF2α and the translational shutoff that the PKR pathway would otherwise induce (reviewed in (Elde and Malik, 2009)). Viruses might also use other strategies that have not yet been characterized to allow viral replication in the face of restriction factors. A key feature common to all these modes of antagonism is the direct interaction between the viral antagonist and the host restriction factor, which has set the stage for the evolutionary arms race that is characteristic of many restriction factors, described below.

Positive selection as a fundamental principle of virus-host interactions

Many non-coding regions of the genome evolve under neutral selection; for example, non-synonymous (amino acid altering) and synonymous mutations are predicted to occur at the same rate in pseudogenes. Most host protein-coding genes evolve under negative (purifying) selection, which removes non-synonymous mutations from the population in order to maintain the function of the protein. By contrast, the interactions between restriction factors and viral antagonists evolve under "positive selection," a selective regime that results in an excess of non-synonymous mutations compared to synonymous mutations. Positive selection is often a result of two genetic entities evolving in conflict with one another, as illustrated by the "Red Queen" hypothesis proposed by Leigh Van Valen to describe an evolutionary system where continuous adaptation is required to maintain the *status quo* (Van Valen, 1973). Virus-host interactions are examples of the "Red Queen" competition, as host restriction factors exert a selective pressure on virus replication and pathogenic viruses exert fitness costs on their hosts. Mutations that allow a restriction factor to evade a viral antagonist provide a means for the host to escape the fitness

costs conferred by the virus. This imposes a selective pressure on the viral antagonist to evolve specificity for a new restriction factor encoded by the host species. As a result, a prey-predator-like 'arms race' dynamic is established, leading to the rapid evolution of both the host and virus (Fig. 2a). Thus, nearly all of the restriction factors described in Table 1 contain signatures of positive selection.

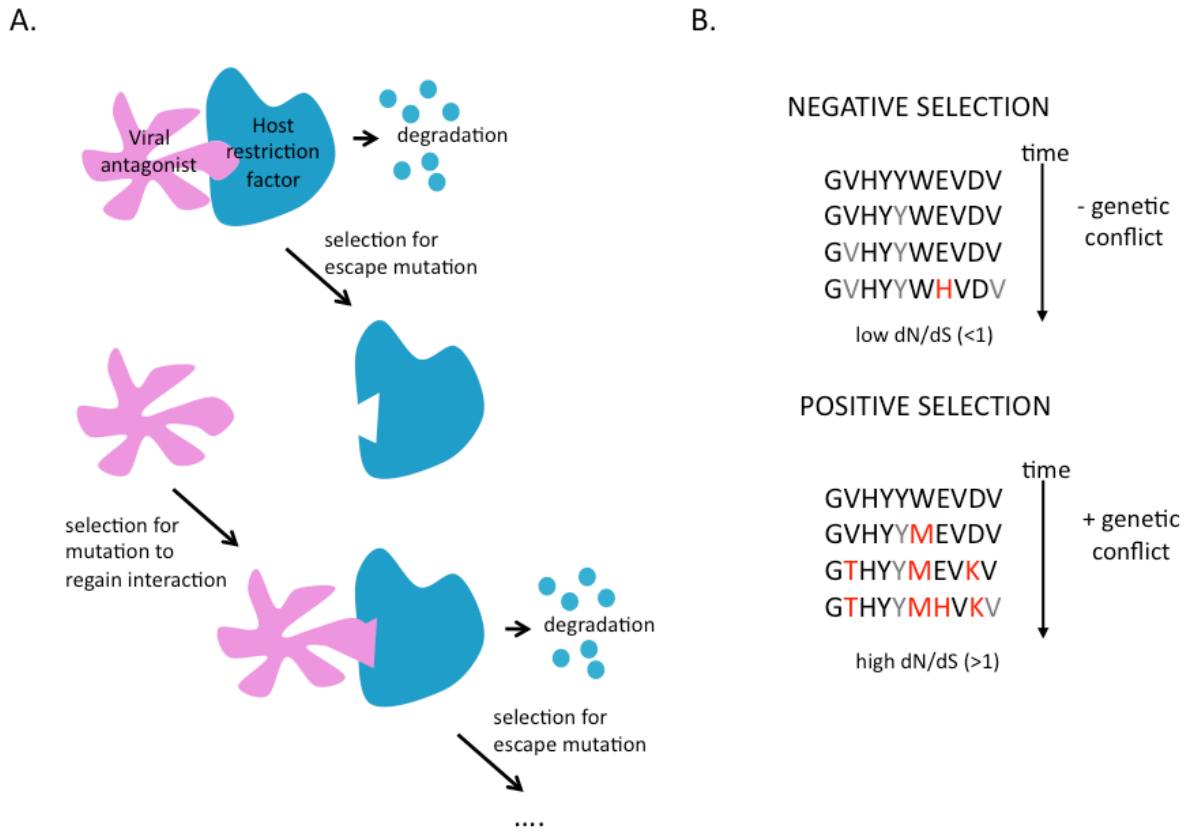


Fig. 2 Genetic conflict between virus and host.
(A) A host restriction factor (blue) that is antagonized by a viral factor (pink) cannot restrict viral replication, and the host is susceptible to viral infection. This exerts a fitness cost on the host, and escape mutations will be selected for in the host factor. In return, when the host factor restricts viral replication, a fitness cost is exerted on the virus. Mutations that allow the virus to re-gain restriction factor antagonism, such as re-forming a protein-protein interface between the viral antagonist and the host factor, are selected for in the virus. This back-and-forth fitness adaptation in the virus and host leads to a conflict that is visible on the genetic level. (B) Over time, in the absence of genetic conflict, most genes evolve under negative (purifying) selection. This leads to a lower rate of non-synonymous mutations (dN) than synonymous mutations (dS) in the host gene, and the dN/dS ratio is predicted to be less than 1. In the presence of a genetic conflict, such as a viral antagonist, the host gene will rapidly accumulate non-synonymous mutations, and the dN/dS ratio is predicted to be greater than 1. Grey letters represent synonymous changes, and red letters represent non-synonymous changes.

Because viruses have existed throughout vertebrate evolution, the arms race between host and virus is ancient (Han and Worobey, 2012). In fact, many host restriction factors have evolved under positive selection for hundreds of millions of years. Under a long-term or recurrent viral selection pressure, a single amino acid in a restriction factor that directly interacts with a viral antagonist may repeatedly mutate many times during evolution, or a restriction factor may accumulate mutations at many different residues in order to escape many different viruses. This will lead to an unusually high ratio of the non-synonymous mutation rate (dN) to synonymous mutation rate (dS) —dN/dS— at single residues or across entire proteins (Fig. 2b). To estimate the dN/dS ratio of a gene, ancestral gene sequences can be reconstructed from orthologous gene sequences from modern-day species that diverged millions of years ago, and statistical methods are used to calculate the rate of evolution across a phylogeny (Nielsen and Yang, 1998). Using this method, many human restriction factors have been found to be evolving under episodic positive selection throughout primate evolution (Table 1).

Selective sweeps are the mechanism of adaptation

An important aspect of positive selection of restriction factors is that the selection (and therefore evolution) of an advantageous mutation acts on a population level. The cost of a viral infection must affect the population's ability to reproduce before it will exert a selective pressure on the population to evolve. During a population-wide infection, some individuals may carry a previously neutral genetic mutation that now confers those individuals with a reproductive advantage in the face of infection, and this advantageous genotype will rapidly rise in frequency until the mutation reaches a frequency of 100%, or fixation. In a classic selective sweep, surrounding regions of the genome are inherited together (known as hitchhiking) with the

genomic sequence that confers the fitness advantage, thus decreasing genetic diversity near the region of the genome under positive selection (reviewed in (Biswas and Akey, 2006; Nielsen, 2005; Nielsen et al., 2007)). Therefore, genomic loci that are under positive selection in a population are predicted to have skewed allele frequencies across long genetic distances surrounding the selected locus. Eventually, if a mutation reaches fixation within a species, between-species comparisons will reveal an excess rate of non-synonymous mutations in this region relative to that expected under neutral selection (Nielsen, 2005).

The Apobec3 family of restriction factors

Apobec3 proteins are cytidine deaminases that block the replication of retroviruses and retrotransposons during reverse transcription. The Apobec3 family is a particularly interesting group of restriction factors because they have many classic hallmarks of restriction factors, such as antiviral activity as their only known function, rapid evolution, and viral antagonists, and also some unique characteristics, such as constitutive expression. Further, Apobec3 proteins have been suggested to play a crucial role in preventing experimental cross-species transmissions of HIV-1 in rhesus macaques (Hatziioannou et al., 2006) (Kamada et al., 2006)) and in controlling murine leukemia virus (MLV) and mouse mammary tumor virus (MMTV) pathogenesis in mice (Okeoma et al., 2007; Takeda et al., 2008). The Apobec3 family of restriction factors is also highly relevant to natural HIV-1 infection. The human Apobec3s most likely to play an *in vivo* role in HIV-1 infection are Apobec3G and Apobec3F because they have anti-HIV activity, are expressed in HIV target cells, and HIV-1 has specifically evolved to antagonize them. Also, HIV-1 isolates contain many G-to-A mutations that are the signature of these Apobec3s.

Mechanism of Apobec3 antiviral activity

Apobec3 restriction of retroviruses requires Apobec3 proteins to be packaged into virions. Apobec3 is expressed in an infected cell (the "producer" cell) and selectively incorporated into an assembling virion (Harris et al., 2003) by binding to the viral protein nucleocapsid (NC) (Schafer et al., 2004). Studies in Apobec3G have shown that binding to NC requires a patch of 4 aromatic residues in its N-terminus (Gooch and Cullen, 2008; Huthoff and Malim, 2007). When Apobec3-containing virus infects a target cell, viral replication is inhibited due to catastrophic hypermutation of viral ssDNA (Harris et al., 2003; Mangeat et al., 2003) or steric interference of reverse transcriptase (RT) processivity (Bishop et al., 2008). Apobec3s induce hypermutation by converting cytidines to uridines in single-stranded DNA (ssDNA), which is an intermediate of reverse transcription. This results in an excess of G-to-A mutations in the provirus. However, some viruses have evolved a mechanism to prevent viral incorporation of Apobec3. The lentiviral protein Vif targets Apobec3 for proteosomal degradation in the producer cell by acting as a substrate receptor in a Cullin5 E3 ubiquitin ligase complex (Mehle et al., 2006; Yu et al., 2003). Vif inhibition of Apobec3 proteins is often species-specific. For instance, human Apobec3G is sensitive to Vif from HIV-1 but not SIV that infects African green monkeys (SIVagm), and Apobec3G from African green monkeys is sensitive to Vif from SIVagm but not HIV-1 (Bogerd et al., 2004; Mangeat et al., 2004; Schrofelbauer et al., 2004; Xu et al., 2004).

Antiviral activity of human Apobec3s

The human Apobec3 proteins have a broad spectrum of strength and targets. For example, human Apobec3A and Apobec3C restrict retrotransposons but not retroviruses (Bogerd et al., 2006a; Muckenfuss et al., 2006; Stenglein and Harris, 2006). Human Apobec3A also inhibits

single-stranded DNA viruses and is highly upregulated by IFN, suggesting that viruses may be its true target (Chen et al., 2006; Koning et al., 2009; Refsland et al., 2010). Apobec3B restricts retrotransposons and HIV-1 but is not expressed in HIV target cells and is resistant to HIV-1 Vif (Doehle et al., 2005; Koning et al., 2009; Refsland et al., 2010). Human Apobec3F restricts HIV-1 and is partially resistant to HIV-1 Vif (Bishop et al., 2004; Liddament et al., 2004; Wiegand et al., 2004; Zheng et al., 2004), but its restriction of retrotransposons is unclear (Hulme et al., 2007; Stenglein and Harris, 2006). Human Apobec3DE moderately inhibits HIV-1 and is sensitive to HIV-1 Vif (Dang et al., 2006), and its restriction of retrotransposons is also unclear (Niewiadomska et al., 2007; Stenglein and Harris, 2006; Tan et al., 2009). Human Apobec3G restricts HIV-1 more potently than any other human Apobec3 (Bishop et al., 2004; Zheng et al., 2004) and restricts LTR- and Alu retrotransposons but not LINE-1 retrotransposons (Esnault et al., 2005; Hulme et al., 2007; Turelli et al., 2004). Human Apobec3H has the broadest activity, as it potently restricts retroviruses and LTR and non-LTR retrotransposons, but its protein expression has not been detected in any tissue type (Li et al., 2010; OhAinle et al., 2008). Human Apobec3DE, Apobec3F, and Apobec3G are thought to be constitutively expressed in broad tissue types, including HIV target cells (Koning et al., 2009; Refsland et al., 2010).

Evolution of the Apobec3 locus

The Apobec3 family of restriction factors is specific to mammals. Rodents have one *Apobec3* gene, suggesting that the ancestral mammalian state was a single *Apobec3* gene (Conticello et al., 2005). Paralogous expansion of the ancestral *Apobec3* gene has led to a family of *Apobec3* genes in primate, artiodactyl, feline, and canine species (Bogerd et al., 2008; Conticello et al., 2005; LaRue et al., 2008; Munk et al., 2008; OhAinle et al., 2006). Humans and rhesus macaque have a

syntenous cluster of seven *Apobec3* genes (Fig. 3) (Hultquist et al., 2011; Jarmuz et al., 2002; OhAinle et al., 2006).

Many Apobec3 genes, including Apobec3G and Apobec3H, have evolved under positive selection for millions of years in primates (OhAinle et al., 2006; Sawyer et al., 2004). At least one Apobec3 gene has acquired polymorphisms within an Old World monkey species that allow host evasion from lentiviral infection in recent history (Compton et al., 2012), suggesting that Apobec3s have likely evolved in recent primate history as well. Recently, several *Apobec3* family members, including *Apobec3B*, *Apobec3G*, and *Apobec3H*, have been shown to be polymorphic in the human population (An et al., 2004; Kidd et al., 2007; OhAinle et al., 2008; Reddy et al., 2010). Population stratification of variants of Apobec3B and H that alter anti-viral activity suggests that the current human *Apobec3* repertoire has recently been under positive selection and is rapidly changing in response to viral pathogens.

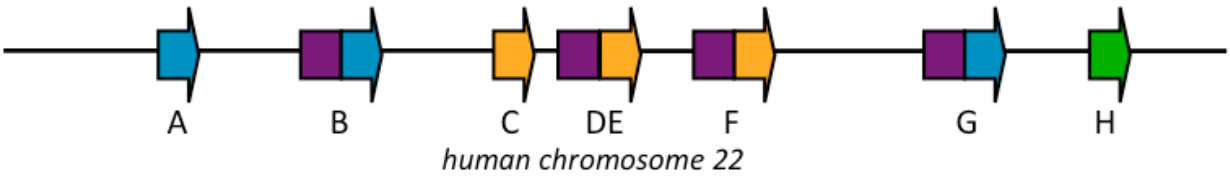


Fig. 3 Expansion of the *Apobec3* locus in primates.
The seven *Apobec3* paralogues are clustered on Chromosome 22 in humans. Colors represent sequence homology.

How hosts keep up in the arms race

If single nucleotide changes were the only effector mechanism in the co-evolution of host and virus, the host would be at a seemingly enormous disadvantage because RNA viruses and some small ssDNA viruses have nucleotide substitution rates that are 1000 times faster than their hosts

(Hanada et al., 2004; Jenkins et al., 2002; Shackelton and Holmes, 2006; Shackelton et al., 2005). How then does a host restriction factor ever win an arms race with a virus, especially when considering that a host might be simultaneously challenged by many different types of viruses? The answer lies in the types of genetic landscape that viruses and hosts can explore.

Limitations on viral evolution

RNA viruses maintain a densely packed genome that includes overlapping reading frames and RNA hairpin structures involved in genome packaging and replication. The limitations on genome size of RNA viruses necessitates that many viral proteins carry out multiple functions (Holmes, 2003). This constrains their evolutionary potential, as mutations to optimize one function (for example, adaptation to a polymorphism in a host restriction factor) might compromise another function of that protein (for example, capsid assembly). In addition, the small viral genome size prevents the use of gene duplication-driven strategies for adaptive evolution, which can be used by host, but not viral, genomes. One possible exception to the lack of gene duplication in viruses is the pair of homologous genes *vpr* and *vpx* in lentiviruses (Sharp et al., 1996).

Host heterozygosity

Genetic polymorphisms in genes encoding host restriction factors are often maintained as a result of population-level adaptation against viruses. Balancing selection in restriction factors may result when multiple viruses co-infect a population, such that different host haplotypes are both advantageous, essentially maintaining polymorphism within the population ('frequency-dependent selection'). Several of the best-known examples of genes under balancing selection

include genes involved in immunity, such as genes in the major histocompatibility complex (MHC) (Hughes and Yeager, 1998), which maintain multiple alleles that present a variety of antigens and therefore protect against a variety of pathogens; and the *glucose-6-phosphate dehydrogenase* (*G6PD*) gene (Verrelli et al., 2002), a housekeeping gene that maintains polymorphisms that are associated with clinical disorders and also with malaria resistance.

Heterozygosity in a restriction factor may be advantageous to a population on a short time scale, as host polymorphisms would force a virus to evolve the ability to target multiple alleles of a given host factor. This was recently suggested for the restriction factor Apobec3G in African green monkeys (Compton et al., 2012), a primate species naturally infected with simian immunodeficiency virus (SIV). Apobec3G is polymorphic in African green monkeys, with some individuals carrying a single amino acid change that renders Apobec3G resistant to its viral antagonist, Vif. In an experimental infection of African green monkeys with SIV, the virus from a monkey that was heterozygous for the Vif-resistant allele was unable to evolve the ability to antagonize Apobec3G , while the virus from a monkey that was homozygous for the Vif-resistant allele was quickly able to evolve the ability to antagonize Apobec3G . This suggests that maintaining polymorphism in a restriction factor can be functionally beneficial.

Gene duplication and innovation

Gene duplication of restriction factors is another evolutionary strategy for accelerating host adaptation to a virus. By duplicating a restriction factor, the host can simultaneously explore multiple evolutionary trajectories. For example, in primate, artiodactyl (cloven-hooved mammal), canine, and feline species, the Apobec3G repertoires include many more paralogues than in rodents. Most primate genomes now encode 7 paralogues of *Apobec3* genes, which vary

in terms of their antiviral activity and retroelement targets, suggesting that they are adapted to different viruses. Several *Apobec3* genes show evidence of positive selection in primates, including *Apobec3G* and *Apobec3H* (OhAinle et al., 2006; Sawyer et al., 2004), but the specific residues under positive selection vary between the *Apobec3* genes, further supporting the idea that each paralogue is evolving against different viral targets. In this way, increasing the copy number of a given restriction factor probably gives the host flexibility to rapidly evolve against several different viruses, leading to large families of related restriction factors. Other restriction factor families that are the result of gene duplications include the *Mx1* gene family with 2 paralogues in some mice (Staeheli and Sutcliffe, 1988), the *IFITM* gene family with at least 4 paralogues in humans and 5 paralogues in mice (Siegrist et al., 2011), and the *Trim5* gene family with 8 paralogues in mice and cows and 3 paralogues in rats (Sawyer et al., 2007; Tareen et al., 2009).

By contrast, multiple members of a restriction factor family could also evolve to target the same virus in different ways, thereby constraining viral evolution such that the virus must maintain multiple defense strategies. An example of this is the pair of human paralogues *Apobec3F* and *Apobec3G*. Apobec3F and Apobec3G deaminate cytidine residues in the viral genome within different preferential sequence contexts (Bishop et al., 2004). Thus, primate lentiviruses such as HIV-1 have had to evolve multiple mechanisms of antagonizing Apobec3 proteins by binding to Apobec3F and Apobec3G using distinct domains of Vif (Russell and Pathak, 2007; Simon et al., 2005; Tian et al., 2006). In this way, the host limits the ability of the virus to evolve while increasing antiviral activity.

Although *TRIM5* is not duplicated in primates, it has undergone gene innovation in macaques and owl monkeys, who have independently gained additional exons through the

insertion of a *Cyclophilin A* (*CYPA*) gene into noncoding segments of the *TRIM5* gene (Brennan et al., 2008; Liao et al., 2007; Newman et al., 2008; Virgen et al., 2008; Wilson et al., 2008). In both cases, a TRIM-CYP fusion protein with potent antiviral activity is produced, though the viral targets are not the same. TRIM-CYP from owl monkeys and TRIM5α restrict HIV-1, and TRIM-CYP from old world monkeys restricts HIV-2. TRIM-CYP and TRIM5α can both be expressed in the same individual, allowing the restriction of multiple lentiviruses.

Just as restriction factor families expand when more antiviral activity is advantageous in the presence of viruses, restriction factor families can also contract in the absence of a selective pressure. For example, the restriction factor *Apobec3H* from macaques and chimpanzees has potent antiviral activity against lentiviruses (OhAinle et al., 2008). However, two independent loss-of-function mutations in *Apobec3H* have risen to high frequency in humans, despite the conservation of antiviral activity in other primates (OhAinle et al., 2008). This suggests that restriction factors such as *Apobec3H* , which are highly active DNA mutators, may impose a cost on the host genome, and they may be selected against in the absence of a viral pressure.

Lessons from studying the co-evolution of restriction factors and viruses

Studying the evolution of restriction factors can help us to understand why humans are susceptible to modern-day viruses, as our immune responses to contemporary viruses have been shaped by our evolutionary responses to previous infections. The modern innate immune system is not yet optimized against modern viruses, but rather was selected for by previous rounds of co-evolution with ancient viruses. By determining what type of viral infections occurred in the past and how they were eliminated, we can form new ideas about how to manipulate the immune system to our advantage in the ongoing battle against viruses.

Identifying previous viral infections

Paleovirology is the study of ancient, extinct viruses (paleoviruses) and their effects on modern day host-virus interactions (Emerman and Malik, 2010). We know that ancient retroviruses infected primates because of remnants of viral sequences found in primate genomes (Katzourakis et al., 2005). However, many retroviruses did not become endogenous in the host genome, and so we have no direct evidence of their existence. In fact, there have been no endogenous lentiviral sequences found in primate genomes other than in a single genus of prosimians (Gifford et al., 2008; Gilbert et al., 2009; Katzourakis and Gifford, 2010). However, by identifying signatures of positive selection in host restriction factors, we can infer the existence of many additional paleoviruses, as well as the historical timeframe and species in which the infection took place. By combining evolutionary analyses with functional tests, we can determine the type of virus that likely drove selection in the host (Fig. 4).

One of the clearest examples of identifying a paleovirus using positive selection comes from the acquisition of a *TRIM-CYP* anti-lentiviral gene fusion in owl monkeys 2-6 million years. This strongly argues for such a challenge occurring in this lineage of primates, which is both phylogenetically and geographically distinct from the primates that are known to be infected with lentiviruses currently. By studying the evolution of restriction factors, we can form a more accurate picture of the ancient history of retroviral infections in primates.

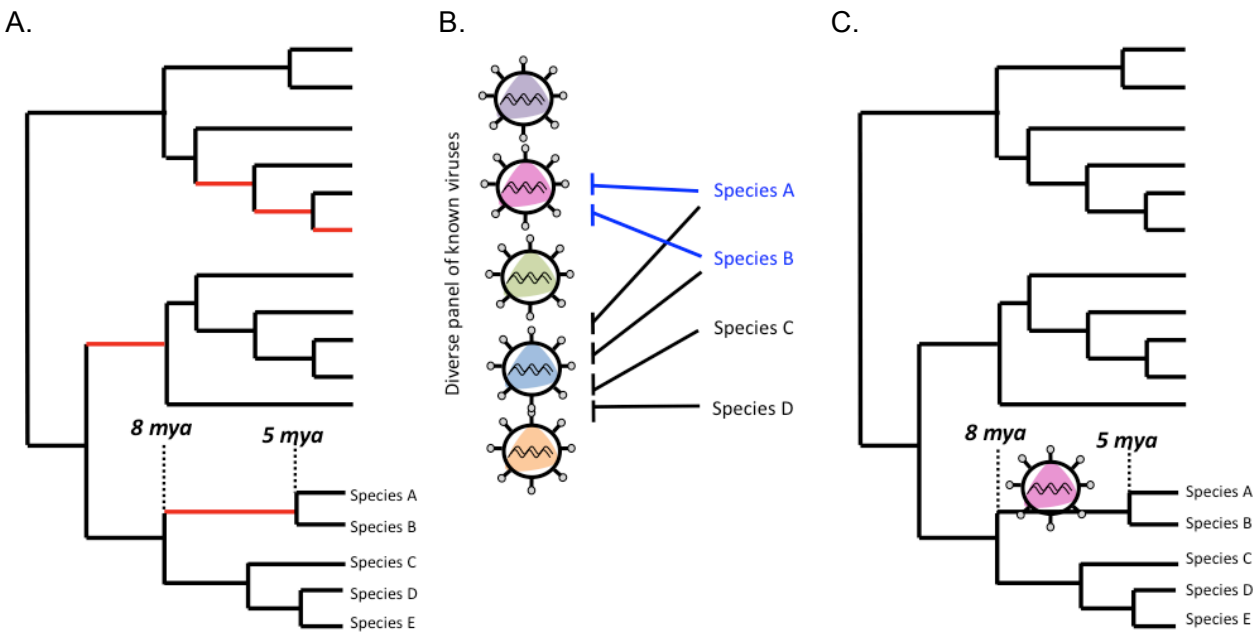


Fig. 4 Identifying paleoviruses using positive selection.
(A) Phylogeny of a restriction factor. Lineages under positive selection (highlighted in red) are identified using likelihood methods of calculating the ratio of the rate of non-synonymous mutations (dN) to the rate of synonymous mutations (dS) for each branch. (B) Infectivity assays. Orthologues of the restriction factors from extant species are tested for antiviral activity against a panel of viral targets, with the goal of finding species-specific antiviral activity (highlighted in blue). (C) Inference of a paleovirus. In lineages that are under positive selection, a paleovirus similar to the virus against which the restriction factor has gained species-specific activity is predicted to have existed during the time of selection.

Predicting viral pathogenicity

In the virus-host arms race, positive selection results when the reproductive fitness of either party is challenged. If a virus is not pathogenic to the host, it is not likely to exert a selective pressure on the host. Therefore, adaptive changes in host restriction factors would not be expected to occur during a non-pathogenic infection. Mildly pathogenic viruses would be expected to impart weak selective pressures that might increase the allele frequency of a selected mutation but would not drive polymorphisms to fixation (Pritchard et al., 2010). Many SIVs that infect old world monkeys and all simian foamy viruses (SFVs) are considered nonpathogenic in their

natural hosts (Meiering and Linial, 2001; Sodora et al., 2009). Interestingly, SFVs have co-evolved with their hosts for 30 million years (Switzer et al., 2005), demonstrating that there might not be a selective pressure to stop SFV replication. Furthermore, the rate of evolution of SFVs is many times slower than for other RNA viruses (Switzer et al., 2005), which suggests that the arms race between virus and host has slowed down considerably in this case. Natural infection of African green monkeys by SIV is also thought to be non-pathogenic, as infection does not cause immunodeficiency despite high viral replication levels (Sodora et al., 2009). Surprisingly, polymorphisms in the African green monkey *Apobec3G* gene that allow evasion from antagonism by host-specific SIV Vifs were found in Grivet and Sabeus subspecies (Compton et al., 2012), suggesting a recent selective pressure on *Apobec3G*. Furthermore, the SIVs that circulate in these subspecies have re-gained the ability to antagonize Apobec3G. This demonstrates that there is an arms race between SIV and African green monkeys that has implications for SIV pathogenesis in African green monkeys. SIV might have been formerly pathogenic to African green monkeys, or its pathogenesis might be present even now in unmeasured or mild ways. In this way, the evolution of a host restriction factor can inform our views of viral pathogenesis.

Coordination with other arms of the immune system

The relationship between restriction factors and the rest of the innate immune system is a growing area of research. In many ways, restriction factors are similar to pattern recognition receptors (PRRs) because they recognize structural patterns on pathogens. In fact, TRIM5, which binds to a viral capsid lattice structure (Ganser-Pornillos et al., 2011) and accelerates capsid uncoating to cause viral restriction, has recently been shown to also function as a PRR for

retroviruses (Pertel et al., 2011). Upon binding to retroviral capsids, Trim5 leads to the activation of nuclear factor-κB (NFκB) signaling and a distinct innate immune response. Moreover, even in the absence of retroviral capsids, Trim5 has been shown to have a role in innate immune responses, as it functions as a constitutive signaling intermediate in the NFκB cascade (Tareen and Emerman, 2011).

SAMDH1 also functions as both a restriction factor and a mediator of the innate immune response to non-viral events. SAMHD1 protects dendritic cells and monocytes from HIV-1 infection by decreasing the level of cellular dNTPs below the level required for synthesis of viral DNA (Hrecka et al., 2011; Laguette et al., 2011; Lahouassa et al., 2012). In addition, SAMHD1 has a role in the innate immune response even in the absence of retroviral infection, as genetic mutations in SAMHD1 are associated with autoimmunity in humans (Rice et al., 2009). Similarly to TREX1, SAMHD1 might have a protective role against autoimmune responses by preventing inappropriate retrotransposon by-products such as single-stranded DNA (Lahouassa et al., 2012; Stetson et al., 2008). Studying retroviral restriction factors will undoubtedly contribute to our understanding of many aspects of the innate immune system.

Explaining why humans are susceptible to modern-day viruses

HIV-1 and HIV-2 are the result of multiple cross-species transmission events of SIV from chimpanzees and sooty mangabeys, respectively, into humans (Sharp and Hahn, 2011). Primate restriction factors have been shown to have an important *in vivo* role in preventing lentiviral cross-species viral transmission events. For example, experimental infection of rhesus macaques – which are not infected with SIV in the wild – with HIV or SIV can mimic a cross-species transmission event. During experimental HIV-1 infection, rhesus macaque TRIM5α and

Apobec3G completely restrict viral replication (Hatziioannou et al., 2006; Kamada et al., 2006). Furthermore, naturally occurring polymorphisms in rhesus macaque *TRIM5* attenuate viral replication by 100- to 1,000-fold during experimental infection with SIV from sooty mangabeys (SIVsm) (Kirmaier et al., 2010). Identification of the host genes involved in susceptibility or resistance to SIV infection may help to explain the dynamics of lentiviral zoonoses.

Our immune system may be better at preventing cross-species viral transmissions than intra-species viral transmissions because viruses that have crossed the species barrier have already partially adapted to the host. Therefore, the evidence for the effect of restriction factors on within-species viral acquisition is less clear. The four groups of HIV-1, which are each the result of an independent cross-species transmission event to humans from chimpanzees infected with SIV (SIVcpz), differ in their global spread, with group M representing the pandemic strain. It has recently been shown that the adaptation of HIV-1 to human-specific mutations in the restriction factor Tetherin was only achieved by group M and N viruses but not the non-pandemic groups O and P (Sauter et al., 2011; Sauter et al., 2009). Clearly Tetherin did not prevent any of the four cross-species transmission events, but it has been suggested that overcoming Tetherin restriction by HIV-1 group M was necessary for efficient replication within humans and pandemic spread (reviewed in (Kirchhoff, 2010)). This intriguing idea may shed light on how restriction factors affect within-species viral transmission. In studies of humans, the effects of restriction factor expression levels and polymorphisms on HIV-1 susceptibility and disease progression have not yielded a consensus viewpoint (reviewed in (Albin and Harris, 2010)).

Identifying host-virus interaction domains and implications for treatment

The interactions between a virus and host restriction factor can be mapped down to distinct protein-protein interfaces and, in some cases, down to single amino acid residues. Because these interaction domains are directly engaged in genetic conflict, they are often the residues that are most rapidly evolving. By looking at genetic signatures of positive selection, the sites involved in protein-protein interactions can be predicted and then tested functionally, as was recently done with remarkable accuracy for SAMHD1(Laguette et al., 2012; Lim et al., 2012). Without knowing anything about the domains of SAMHD1 required for antagonism by the lentiviral protein Vpx, two groups performed positive selection analyses of SAMHD1 using the dN/dS test and identified two different regions of SAMHD1 that have evolved very rapidly in primates. When functionally tested, these two regions of SAMHD1 were both shown to confer virus-specific degradation by Vpx from different lentiviruses. This information has helped to explain how lentiviruses and primates have co-evolved. By mapping host-virus interactions, the constraints of the evolutionary arms race can be more fully understood.

Moreover, these protein-protein interactions between host restriction factors and viral antagonists provide tempting targets for small molecule inhibitors. An ideal inhibitor of a viral antagonist would specifically disrupt the ability of the virus to bind the host restriction factor or other host machinery required for antagonism. This would allow a host restriction factor to specifically inhibit viral replication without altering the rest of the immune system of an individual. Inhibitors of viral antagonists could be used as therapeutic treatment in combination with other anti-retroviral drugs. Several inhibitors of HIV-1 Vif have been identified (Cen et al., 2010; Nathans et al., 2008), and attempts at disrupting Vpu function have been made (Zhang et al., 2011). Achieving an inhibitor of a viral antagonist without disrupting host functions might be

difficult because many viral antagonists use or mimic host machinery for their activity. Also, the virus might be able to quickly evolve resistance mutations, as genes encoding viral antagonists often do not have as many functional constraints as more conserved viral genes.

Restriction factors are early, potent, and specific cellular blocks against retroviral replication. They have clearly had an important role in innate immunity throughout primate evolution, and characterizing the evolution of restriction factor antiviral activity will help us to understand why we are winning or losing current battles against viruses. Further, we can identify potential host restriction factors by their characteristics: antiviral activity, IFN induction, viral antagonism, and rapid evolution. In previous studies, Apobec3A, B, C, F, G, and H have been described in detail with regards to these characteristics. However, a clear role in antiviral host defense has not yet been identified for Apobec3DE, despite having the greatest sequence divergence between humans and chimpanzees of all Apobec3 genes and a known viral antagonist. Therefore, I was interested in understanding the potential role of primate Apobec3DE in host defense.

In this work, I will describe my efforts to define the evolution and antiviral activity of Apobec3DE in primates (Chapter III) and of the Apobec3 locus in human populations (Chapter IV). I will also discuss the implications of this work for our understanding of the susceptibility of humans to retroviral infection and, more generally, to our understanding of host-virus interactions (Chapter V).

Chapter II
Materials and methods

Apobec3 sequences

Human Apobec3DE was amplified from a testes cDNA library (Invitrogen). This sequence is identical to the sequence in the GenBank database (NM_152426.3). Chimpanzee (*pan troglodytes*) Apobec3DE was amplified by reverse transcriptase (RT)-PCR of RNA isolated from fibroblasts (Coriell AG06939) followed by nested PCR. The PCR fragment was subcloned into a TOPO vector (Invitrogen) for sequencing. This sequence differs from the NCBI database entry for the predicted chimpanzee sequence (XM_525657.2) across the entire first exon. The sequences for bonobo (*pan paniscus*) *Apobec3DE* exons 1-7 and gibbon (*nomascus leucogenys*) *Apobec3DE* exons 1-2 were obtained by amplifying genomic DNA isolated from fibroblasts (Coriell AG05253 and PR01037) using hominoid-specific Apobec3DE genomic DNA primers. The sequences for rhesus macaque (*macaca mulatta*), African green monkey (*cercopithecus aethiops*), and Patas monkey (*erythrocebus patas*) *Apobec3DE* exons 1-7 were obtained by amplifying genomic DNA isolated from fibroblasts (Coriell AG07098 and AG06254) or cell lines (Vero) using Old World monkey-specific Apobec3DE genomic DNA primers. PCR products were sequenced directly. The sequence for rhesus macaque *Apobec3DE* differs from the GenBank entry (XM_002798351.1) by 7 nucleotides. The sequences for gorilla (*gorilla gorilla*) Apobec3DE and gibbon Apobec3DE (exons 2-7) were amplified from RNA isolated from fibroblasts (Coriell NG05251) using 3'RACE followed by nested PCR and subcloned into the pGEM-T Easy vector (Promega) for sequencing. cDNA was synthesized using the oligo $(dT)_{17}$-Adaptor primer described in (Sambrook and Russell, 2006). Gorilla Apobec3F was amplified from RNA by One-Step RT-PCR (Qiagen). All primer sequences are in Table 2. Sequences have

been deposited in GenBank under the following accession numbers: JN247640-JN247649. Additional sequences for Apobec3F were obtained from the GenBank database: human Apobec3F (NM_145298.5), chimpanzee Apobec3F (XM_525658.2), and rhesus macaque Apobec3F (NM_001042373.1). The sequence for African green monkey Apobec3F was provided by Vinay Pathak.

Primate Apobec3 sequence analysis

Apobec3 sequences were analyzed with the Phylogeny.fr web service (Dereeper et al., 2008) using default settings, including maximum likelihood phylogeny construction with an Approximate Likelihood-Ratio Test (aLRT) for branch support. Recombination breakpoints were identified using the GARD method with an HKY95 nucleotide substitution model on the Datamonkey web server (Kosakovsky Pond et al., 2006). Evolutionary analyses of the phylogeny were performed with the CODEML program in the PAML software package (Yang, 2007). Global dN/dS ratios for the branches were calculated using a free-ratio model. We calculated the log-likelihood ratios of data under different NSsites models (Model 1 vs. Model 2 and Model 7 vs. Model 8). For both cases, a model of positive selection gave a better fit ($p < 0.001$). A Bayes empirical bayes analysis identified amino acid residues with a high posterior probability (>0.95) of evolving under positive selection (Yang et al., 2005).

Human Apobec3 genotype data

SNVs in the Apobec3 locus (GRCh37, Chr22: 39347756 – 39501072) from 1094 individuals were obtained through the 1000 Genome Project May 2011 phase 1 low-coverage phased genotype release. Insertions and deletions were obtained for 1092 individuals from the 1000

Genome Project February 2012 integrated phase 1 release. The African population includes 61 African ancestry individuals from southwest US, 97 Luhya individuals in Webuye, Kenya, and 88 Yoruba individuals from Ibadan, Nigeria. The Asian population includes 97 Han Chinese individuals in Beijing, 100 Chinese individuals in Denver, and 89 Japanese individuals in Tokyo. The European population includes 87 individuals with European ancestry in Utah, 93 Finnish individuals from Finland, 89 British individuals from England and Scotland, 14 individuals from Iberian populations in Spain, and 98 Toscani in Italy. The South American population includes 60 Colombian individuals in Medellin, Colombia, 66 Mexican individuals from LA California, and 55 Puerto Rican individuals in Puerto Rico. The ancestral state of each allele was determined by comparison to the chimpanzee genome.

Confirmation of rare Apobec3 variants

Human B cell lines from individuals participating in the 1000 Genome Project were obtained from Coriell (GM18963, GM19655, and HG01102). RNA and genomic DNA were isolated from B cell lines (Qiagen). Apobec3G was amplified from B cells RNA by One-Step RT-PCR (Qiagen) using forward and reverse Apobec3G primers followed by semi-nested PCR using the same forward primer and a nested Apobec3G reverse primer (Table 2). The deletion of Apobec3F was confirmed by PCR on genomic DNA from B cells using forward and reverse Apobec3F genomic deletion outside primers (Table 2). As a control, a region within the deletion was amplified using the same forward primer and the Apobec3F genomic deletion inside reverse primer (Table 2).

Population genetics analyses

Population genetics calculations were performed using DnaSP v5 (Librado and Rozas, 2009). Significance of the McDonald Kreitman test was determined using Fisher's exact test of independence. The HKA test was performed using by comparing Apobec3 genes to a 2 kb noncoding region of the Apobec3 locus (Chr. 22: 39403000 – 39405000). Significance of the HKA test was determined using the χ^2 value with 1 degree of freedom. Significance of Tajima's D and Fay and Wu's H tests were determined using 10,000 coalescent simulations based on the observed θ with no recombination. The outgroup used for all tests was the UCSC chimpanzee sequence, accessed through the SeattleSeq Annotation 134 webserver. Sliding windows were calculated with a window size of 100 nucleotides and an overlap of 25 nucleotides. Linkage disequilibrium (r^2) was calculated using Haploview v4.2.

Apobec3 plasmids

Human Apobec3A, Apobec3C, Apobec3DE, and Apobec3G were cloned into the expression vector pCS2, and human Apobec3F was cloned into the expression vector pCDNA3.1. A hemagglutinin (HA) tag was added to the N-terminus of Apobec3A and Apobec3G in Chapter III and to the C-terminus of all other Apobec3 proteins in Chapters III and IV. Untagged A3DE constructs were also made. A Kozak sequence was added upstream of each start codon, except for Apobec3DE constructs. Human/chimpanzee 1 and 2 chimeras were made by ligating BamHI fragments. Human/chimpanzee 3 and 4 chimeras and all point mutants for Apobec3DE, Apobec3A, Apobec3C, Apobec3F, and Apobec3G were constructed by overlapping PCR products.

Cell lines, Transfections, and Western Blot analysis

293T, CRFK, HeLa, and HeLa-HA cells were maintained in DMEM/1% Pen/Strep/10% BGS at 37^{o}C in a CO_2 incubator. B cell lines, SupT1 cells, H9 cells, and human PBMCs were maintained similarly in RPMI/1% Pen/Strep/10% FBS. Transfections were performed with TransIT-LT1 transfection reagent (Mirus Bio) at a reagent:plasmid DNA ratio of 2:1. Cell and viral lysates were harvested 48 hours after transfection. Cells were lysed in NP40-doc buffer (1% NP40, 0.2% sodiumdeoxycholate, 0.12M NaCl, 20mM Tris pH 8.0, 2.4mM DTT) with protease inhibitors (Roche) and spun for 1 minute at 16,000xg, and supernatants were quantified by Bradford assay. Virus was quantified by p24 gag ELISA (Advanced BioScience), and 20 ng virions were spun down at 16,000xg for 45 minutes and lysed in protein loading dye (125mM Tris pH 6.8, 4% SDS, 20% glycerol, 10% β-mercaptoethanol, 6M Urea, bromophenol blue). Lysates were resolved by 10% SDS-PAGE, transferred to PVDF membranes, and probed with anti-HA (Santa Cruz Biotech), anti-actin (Sigma) antibodies, or anti-Apobec3G antibodies. Relative protein levels were quantified using ImageJ. Of note, lysates containing human Apobec3DE that were prepared without Urea did not fully denature the protein, preventing its appropriate migration through the gel. This requirement for Urea in protein preparations was not observed for chimpanzee Apobec3DE.

Retrotransposition assays

MusD, LINE-1, and Alu assays were performed as described previously (Dewannieux et al., 2003; OhAinle et al., 2008; Ribet et al., 2004; Wei et al., 2001) with some modifications. Cells were plated in 6-well plates at 8x10^4 cells/well, and, 24 hours later, cells were transfected. For MusD assays, 0.25 μg *neo*^{TNF} marked MusD plasmid (Ribet et al., 2004) and 0.6 μg Apobec or

empty plasmid were transfected into HeLa cells. For L1 assays, 0.1 μg LINE-1 plasmid (JM101/L1.3 (Moran et al., 1999; Wei et al., 2001)) and 0.6 μg Apobec or empty plasmid were transfected into HeLa cells. For Alu assays, 1 μg Alu plasmid (Alu*neo*^{TET} (Dewannieux et al., 2003) or Alu-Ya5-eab, Alu-Yc1-eab, Alu-Y-eab, Alu-Sx-eab, and Alu-Jo-eab (Bennett et al., 2008)), 0.3 μg Orf2p (pCep5'UTRORF2Δ*neo* (Alisch et al., 2006)), and 0.45 μg Apobec or empty plasmid were transfected into HeLa-HA cells. Three days after transfection, cells were selected in G418 for 10-12 days. For MusD and L1 assays, cells were split into 10cm^2 plates at the time of selection into multiple dilutions. Colonies were stained with crystal violet and counted manually.

Viral infectivity assays

Single-round MLV, HIV-1, HIV-2, and SIVagm infectivity assays were performed as described (Yamashita and Emerman, 2004). To produce VSV-G-pseudotyped MLV, 4.2x10^6 293T cells were plated in a 12-well plate, and, 24 hours later, co-transfected with 0.5 μg *gag-pol* expression plasmids (pCS2-mGP (Yamashita and Emerman, 2004)), 0.1 μg tat (pCMVtat), 0.5 μg murine retrovirus-based vector expressing luciferase (pLNC-luc (Yamashita and Emerman, 2004)), 0.25 μg L-VSV-G (Bartz and Vodicka, 1997), and 0.5 μg Apobec or empty plasmid. To produce VSV-G-pseudotyped HIV-1, HIV-2, and SIVagm, 3x10^6 293T cells were plated in a 6-well plate, and 24 hours later, co-transfected with 0.6 μg lentiviral vector (pLai3Δ*env*Luc2 (Yamashita and Emerman, 2004), pLai3Δ*env*Luc2Δ*vif* (OhAinle et al., 2006), pROD9Δ*env*EGFP, pRODΔ*env*EGFPΔ*vif*, pSIVagmTANΔ*env*Δ*vpr*Luc (Mariani et al., 2003), pSIVagmTANΔ*env*Δ*vpr*LucΔ*vif* (OhAinle et al., 2006), 0.2 μg L-VSV-G, and 0.2 μg Apobec or empty plasmid. HIV-1 expressing Vif from SIVcpz was made using

pLai3Δ*env*Luc2Δ*vif*Lk/vif$_{SIVcpz}$ (Li et al., 2010), and HIV-1 expressing Vif from HIV-2 was made using pLai3Δ*env*Luc2Δ*vif*Lk/vif$_{ROD9}$ (Li et al., 2010). For HIV-1 assays, media was changed to DMEM/1% FBS 24 hours before harvesting virions. All viruses were harvested 48 hours after transfection and filtered through a 0.2μm filter. For MLV and HIV-2 infections, equivalent volumes of virus were used to infect $8x10^5$ CRFK cells or $1x10^6$ SupT1 cells, respectively, in a 12-well plate in the presence of 20 ug/mL DEAE-dextran, followed by a 30-minute spinoculation at room temperature. For HIV-1 and SIVagm infections, virus equivalent to two nanograms of p24 gag or four nanograms p27 gag, respectively, was used to infect $4x10^6$ SupT1 cells in a 96-well plate in the presence of 20 ug/mL DEAE-dextran. Forty-eight hours after transfection, cells from triplicate infections were lysed in 100 uL of Cell Culture Lysis Buffer (Promega) or Bright-Glo Luciferase Assay Reagent (Promega) and read on a luminometer.

Pulse-chase analysis

Pulse-chase and immunoprecipitatons was performed essentially as described (OhAinle et al., 2008). Briefly, $1.25x10^6$ 293T cells were plated on $6cm^2$ poly-L-lysine plates (Becton Dickinson) and transfected with 5 μg human or chimpanzee Apobec3DE plasmids. Cells were pulsed with 200 μCi EasyTag EXPRE^{35}S^{35}S Protein label (PerkinElmer) per plate and chased for 0 minutes, 30 minutes, 1 hour, and 20 hours. Cells were scraped into 500μl NP40-doc lysis buffer, and lysates were spun for 1 minute at 16,000xg. Lysates were pre-cleared with protein G-sepharose beads, and Apobec3DE proteins were immunoprecipitated with beads pre-conjugated to anti-HA antibody (Santa Cruz Biotech). Immunoprecipitated proteins were resolved by 10% SDS-PAGE, and the gel was imaged using a Typhoon Trio imager (Amersham Biosciences).

Detection of G-to-A hypermutation

VSV-G pseudotyped HIV-1Δ*vif* containing Apobec3 was produced in 293T cells. Twenty-four nanograms of virus were used to infect $5x10^5$ SupT1 cells in 6-well plates. Twenty-four hours after infection, extra-chromosomal DNA was harvested using a plasmid isolation kit (Roche). DNA was treated with DpnI for 1 hour at 37°C to eliminate plasmid DNA carry over from the initial transfection. A 431-bp fragment of HIV-1 pol was amplified using the previously described forward primer pol1721 and reverse primer pol2152 (OhAinle et al., 2006), subcloned into the pGEM-T Easy vector (Promega), and sequenced.

Deaminase assay

A gel-based deaminase activity assay using cell lysates was performed as described (Kaiser and Emerman, 2006; Thielen et al., 2007) with modifications. Lysates were prepared by sonication in HE buffer (10 mM HEPES [pH7.4], 1 mM EDTA, 1 mM dithiothreitol) with protease inhibitors. Lysates were incubated with assay buffer containing 50 mM Tris [pH7.4], 10 mM EDTA, 0.1 pmol IR700-labeled oligonucleotide, 250 μg/mL RNase A (Qiagen), and 0.4 units uracil DNA glycosylase (NEB) at 37° C overnight. Reactions were incubated in stop buffer at 95° C for 15 minutes and electrophoresed on a 20% polyacrylamide TBE-urea gel. The IR700 tag was detected by a LI-COR infrared scanner (LI-COR Biosciences).

Velocity sedimentation

Apobec3 complexes were analyzed essentially as described (Thielen et al., 2007). Briefly, cells were prepared in NP40 buffer (0.1 M Tris [pH7.4], 0.1 M NaCl, 10 mM EDTA, 0.625% NP40, 50 mM KAc) with protease inhibitors and lysed with a 20-gauge catheter. Some lysates were

treated with 250 µg/mL RNase A for 1 hour at 37° C. Lysates were layered onto step gradients containing 900 µl each of 15%, 20%, 30%, 50%, and 60% sucrose in NP40 buffer and subjected to velocity sedimentation in a Beckman SW55Ti rotor at 45,000 rpm for 37 minutes at 4° C. Fractions of 450 µl were collected from the top of the gradient, and aliquots were analyzed by Western blot.

Chimp Apobec3DE RNA primers		**5' to 3' sequence**
First round	F	GGGAGACTGGGCCAAGCG
First round	R	GGAGACAGACCATGAGGCCC
Second round	F	TCTAAGAGGCTGAACATGAATC
Second round	R	AGACCCCTCACTGGAGAATCTC
Hominoid genomic Apobec3DE DNA primers		
Exon 1	F	CAGCACTTCAAAAAAAGAGGGAGACT
Exon 2	R	CGCCGCCAGACATTTTACTTATGT
Exon 3	F	CAGTACCATAGCAATTAACACTGAAGGTC
Exon 4	R	CAGGGGGGAGGCACCA
Exon 5 (v2)	F	CCTGCTGCCCCTGCCAGTAC
Exon 5 (v2)	R	ACGCGGGGCACGTCACGTA
Exon 6	F	AATAAATAAGAAAATGCCCTGGGGCTT
Exon 7	R	ATGGAGGGGAGGCCCGTG
Old World Monkey Apobec3DE genomic DNA primers		
Exon 1	F	CAGCGCTTCAAAAAAAAGAGCGGG
Exon 2	R	CCACCGGGCATTTTACTTACGTG
Exon 3	F	AGTCACTGCTGCTCCATGACACCA
Exon 4	R	GGGGGTGGATGGAGGTCG
Exon 5	F	CCCTGCCAGTATCCCCAGTAT
Exon 5	R	CAAGAGGCGGAGGTTGCCTG
Exon 6	F	CTTGGGAGGTCGAGGCTGC
Exon 7	R	TGAGGCAGGGGAGCACAGTG
Apobec3DE 3' RACE primers		
$(dT)_{17}$ -Adaptor	R	GACTCGAGTCGACATCGA$(T)_{17}$
Adaptor	R	GACTCGAGTCGACATCG
Gorilla exon 1	F	ACAGCACTTCAAAAAAAGAGGGAGACT
Gorilla exon 2	F	GTCGAATCACAGGCAGGAGGTG
Gorilla exon 4	R	AGCGTGTGGTGCAGGGATGC
Gorilla exon 7	R	GAGGCCCAGGGAGACCCC
Gibbon exon 2	F	AATCGAATCACAGGCAGGAGGTG
One step RT-PCR primers		
Gorilla Apobec3F	F	GTGAAACCCTGGTGCTCCAGAC
Gorilla Apobec3F	R	GAGGCGGTGAGGCAGGGGA
Human Apobec3G	F	GTGCTCCAGACAAAGATCTAGTCG
Human Apobec3G	R	CTGCCTTCCTTAGAGACTGAGG
Human Apobec3G nested	R	GCAGGGGCTCCAGGATATGTAC
Human Apobec3F genomic DNA primers		
Human Apobec3F outside deletion	F	GCCAACATGGTGAAAACCCAC
Human Apobec3F outside deletion	R	GCACCCCCACCACCTTGG
Human Apobec3F inside deletion	R	CTGCCTCAGTCTCCCAGAATG

Table 2. Primer sequences.
Forward (F) or reverse (R) primer sequences, written 5' to 3'.

Chapter III

Positive selection of Apobec3DE in chimpanzees has driven breadth in antiviral activity

Introduction

Genetic conflict between primates and the pathogens that infect them has led to the rapid evolution of many host defense factors, including the *Apobec3* gene family (Sawyer et al., 2004). Several members of the *Apobec3* family, including *Apobec3B, C, DE, F*, *G*, and *H*, are highly divergent between humans and chimpanzees (OhAinle et al., 2006; Sawyer et al., 2004). The *Apobec3* gene with the highest human-chimpanzee divergence is *Apobec3DE* (Sawyer et al., 2004), but analyses to date have not determined whether this rapid evolution occurred in the human lineage, the chimpanzee lineage, or both. Previous studies have shown that human Apobec3DE has weak activity against HIV-1 and other retroviruses (Dang et al., 2006; Smith and Pathak, 2010) and moderately restricts the non-autonomous Alu retrotransposon (Tan et al., 2009). There are conflicting reports about whether or not Apobec3DE restricts the non-LTR retrotransposon LINE-1 (Niewiadomska et al., 2007; Stenglein and Harris, 2006). In addition to weak activity, human Apobec3DE has been reported to have relatively low protein expression compared to other Apobec3 proteins during transient transfection (Dang et al., 2006; Niewiadomska et al., 2007; Tan et al., 2009). However, many studies have excluded Apobec3DE from their assays due to confusion surrounding its initial classification as two separate genes (*Apobec3D* and *Apobec3E*), one of which was incorrectly assigned as a pseudogene due to

sequencing errors (Jarmuz et al., 2002). Thus, despite its strong signal of adaptive evolution, a clear role for human Apobec3DE in host defense has not yet been identified.

Viral restriction by Apobec3s can be overcome by lentiviruses encoding a Vif protein, such as HIV-1, by causing the degradation of Apobec3s in producer cells via the proteosome (Malim and Emerman, 2008). The sensitivity of Apobec3 proteins to Vif antagonism has been shown to be species-specific. However, Apobec3 specificities in modern primates were more likely to have been driven by genetic "arms races" against ancient viruses (called paleoviruses) rather than against modern viruses (Emerman and Malik, 2010; Sawyer et al., 2004). Thus, while human Apobec3DE has weak antiviral activity against HIV-1, it may restrict ancient retroelements, and another lineage of Apobec3DE may have an entirely different specificity. For this reason, and because human and chimpanzee *Apobec3DE* sequences are very divergent, I wished to determine whether orthologs of Apobec3DE may have differential antiviral activities.

Here, I discover that human and chimpanzee orthologs of Apobec3DE share the ability to inhibit a broad range of retrotransposons, including Alu elements, suggesting that they may play a current role in host defense against these elements. Strikingly, chimpanzee Apobec3DE can also potently restrict some lentiviruses, including HIV-1 and SIVagm, but not HIV-2. On the other hand, human Apobec3DE has only poor activity against HIV-1 and SIVagm. I find that, despite being highly incorporated into virions, human Apobec3DE lacks the ability to hypermutate HIV-1 genomes, while chimpanzee Apobec3DE induces high levels of hypermutation in HIV-1. Through evolutionary analyses, I find that chimpanzee Apobec3DE evolved very rapidly in chimpanzee-bonobo ancestors, including within the domain responsible for differential antiviral activity. The specific adaptation of chimpanzee Apobec3DE to lentiviruses together with its rapid evolution approximately 2-6 million years ago suggests that

Apobec3DE was selected to acquire new functions in the defense of chimpanzees against an ancient retroelement.

Results

Characterization of transfected Apobec3DE protein expression

Because the weak anti-HIV activity of human Apobec3DE might be due to low protein expression, we investigated the predicted start codon. We found that the 5' end of the human *Apobec3DE* gene has two potential start sites, and the start site used in previous studies is located within a poor context for translation. An in-frame ATG eight amino acids downstream contains a better Kozak sequence, which might be used for endogenous translation. We sought to determine whether the correct start codon had been used in previous studies.

We cloned the cDNA for chimpanzee *Apobec3DE* from fibroblasts, which included both potential start sites. Frameshift (FS) mutants of human and chimpanzee Apobec3DE were made by deleting a single nucleotide between the first and second potential start sites (Fig. 5a). These FS mutants only translate protein using the second potential start site. Deletion (DEL) mutants of human and chimpanzee Apobec3DE were made by removing the second potential start site. These DEL mutants only translate protein from the first potential start site.

I expressed wild-type (WT), FS, and DEL human and chimpanzee Apobec3DE proteins in 293T cells. Cell lysates were analyzed by Western Blot against a C-terminal HA tag (Fig. 5a). WT and DEL human and chimpanzee Apobec3DE proteins had equal expression levels, where as FS mutant proteins had no detectable expression. This confirms that the first start site of Apobec3DE is used preferentially by endogenous translation machinery and should be included in the ORF.

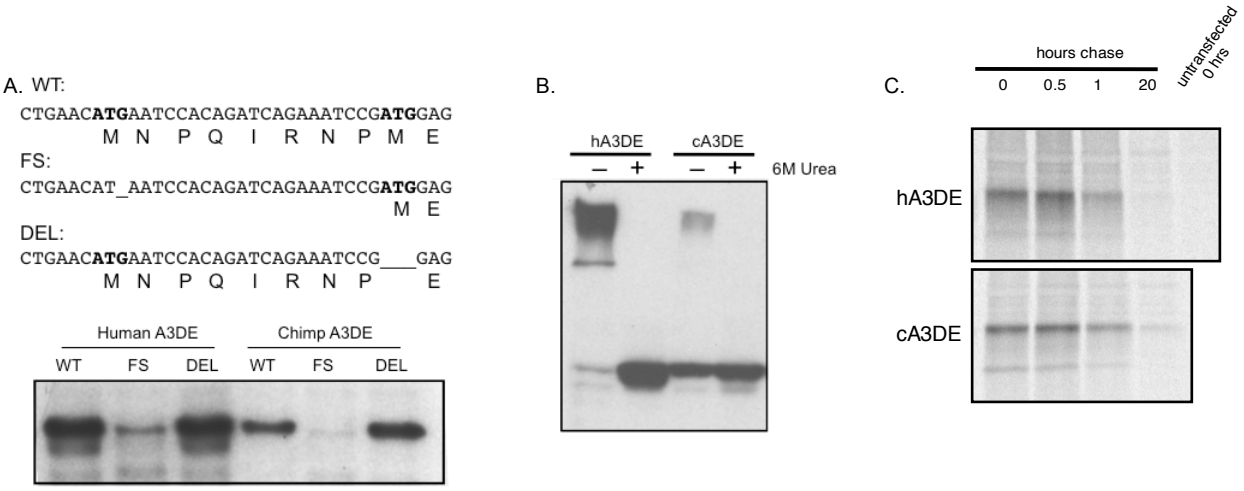


Fig. 5. Expression of transfected Apobec3DE.
(A) (Top) Nucleotide and protein sequences Apobec3DE mutants (WT, FS, DEL). Human and chimpanzee Apobec3DE are identical at these positions. Potential start codons are in bold. Predicted protein sequence is below each nucleotide sequence. (Bottom) Western blot analysis of WT, FS, and DEL mutants of human and chimpanzee Apobec3DE. (B) Western blot analysis of Apobec3DE. Lysates from 293T cells transfected with human Apobec3DE (hA3DE) or chimpanzee Apobec3DE (cA3DE) were prepared in loading dye with or without 6M Urea. The lower band represents the correct molecular weight of Apobec3DE. The upper band is within the stacking gel. (C) Pulse-chase analysis of Apobec3DE proteins. Human and chimpanzee Apobec3DE were radiolabeled for 30 minutes and chased for 0 minutes, 30 minutes, 1 hour, and 20 hours. Proteins were immunoprecipitated using an anti-HA antibody and resolved by SDS-PAGE. Autoradiography was used to visualize the gel.

Next, I tested whether human Apobec3DE protein might be unstable, leading to low cellular protein expression, a characteristic of several haplotypes of human Apobec3H (OhAinle et al., 2008). I performed pulse-chase analyses of human and chimpanzee Apobec3DE by radio-labeling overexpressed proteins in 293T cells. I followed the stability of the proteins out to 24 hours and found that human Apobec3DE has a half-life similar to that of chimpanzee Apobec3DE (Fig. 5c).

However, I did see only low levels of human Apobec3DE at the correct molecular weight by Western blot (Fig. 5b, lower band), as well as a large amount of a high molecular weight complex (upper bands). When lysates were prepared by adding 6M urea to the protein loading

dye, I was able to completely denature the aggregated human Apobec3DE protein. I observed a similar phenomenon with chimpanzee Apobec3DE protein, but to a much lesser extent. From these data, I can conclude that the intrinsic stability of the human Apobec3DE protein is intact and is not responsible for its weak restriction of HIV-1.

Human and chimpanzee Apobec3DE restrict retrotransposons

The sequences for human and chimpanzee Apobec3DE are very divergent (Sawyer et al., 2004) and differ by 35 amino acids, suggesting they may have differential antiviral activity. Previous reports on the activity of human Apobec3DE against retrotransposons have varied, suggesting it could have weak or strong effects (Niewiadomska et al., 2007; Stenglein and Harris, 2006; Tan et al., 2009). No previous study has examined chimpanzee Apobec3DE. Therefore, I characterized the breadth of Apobec3DE restriction by testing the ability of human and chimpanzee Apobec3DE to inhibit several types of retrotransposons. I co-transfected Apobec3DE plasmids with LTR and non-LTR retrotransposons that express neomycin resistance after retrotransposition (Dewannieux et al., 2003; Ribet et al., 2004; Wei et al., 2001). After selection, neomycin resistant colonies were counted as a measure of retrotransposition. Human Apobec3G or Apobec3A were used as positive controls in these experiments, as they have previously been shown to restrict retrotransposons. A western blot shows similar expression levels of human Apobec3G, human Apobec3A, human Apobec3DE, and chimpanzee Apobec3DE when transiently transfected (Fig. 6a).

I found that human and chimpanzee Apobec3DE are strong inhibitors of specific retrotransposons. The most potent activity of human and chimpanzee Apobec3DE was against the LTR retrotransposon MusD (25- and 175-fold restriction, respectively, Fig. 6b). Also, both

human and chimpanzee Apobec3DE restricted the prototypic Alu element cloned by Dewannieaux *et al* (Dewannieux et al., 2003) by 35- and 25-fold, respectively (Fig. 6c). However, neither chimpanzee nor human Apobec3DE had a strong effect on the non-LTR retrotransposon LINE-1 (3-fold, Fig. 6d), as previously shown for human Apobec3DE (Stenglein and Harris, 2006).

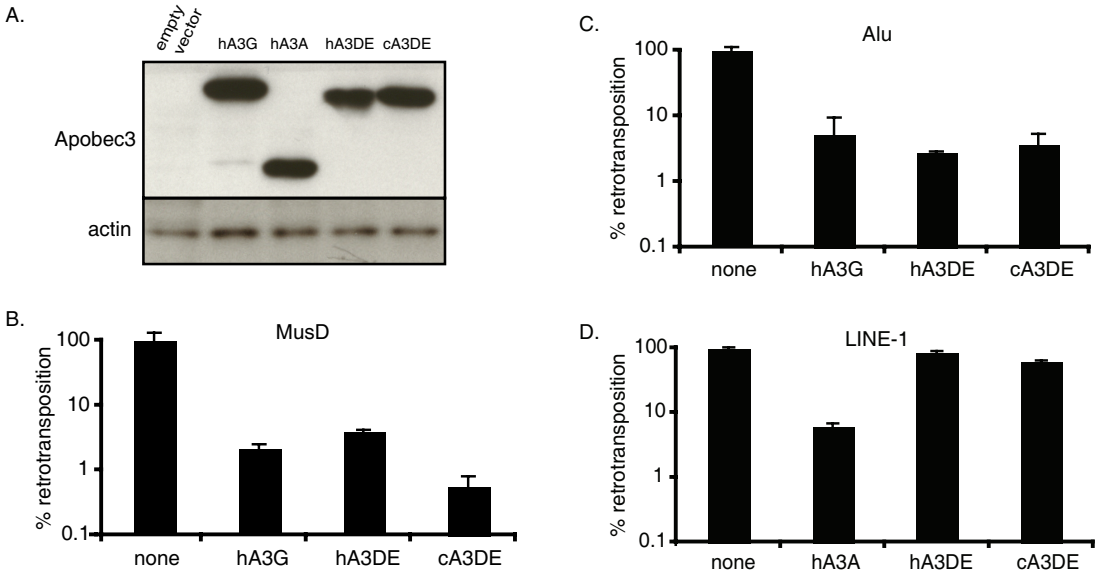


Fig. 6. Human and chimpanzee Apobec3DE restrict retrotransposons.
Human and chimpanzee Apobec3DE were expressed in retrotransposition assays to determine their antiviral activity. Retrotransposition activity is normalized to 100% in the absence of Apobec3 (none). (A) Western blot analysis of protein levels of human Apobec3G (hA3G), human Apobec3A (hA3A), human Apobec3DE (hA3DE), and chimpanzee Apobec3DE (cA3DE) during transient transfection. (B) Retrotransposition of MusD. Human Apobec3A was used as a positive control. (C) Retrotransposition of the prototypic Alu element. Human Apobec3G was used as a positive control. (D) Retrotransposition of LINE-1. Human Apobec3G was used as a positive control. Experiments were performed at least 3 times, and results from one representative experiment are shown. Error bars represent the standard deviation of triplicate transfections within one experiment.

Because different Alu subfamilies have been active in humans compared to chimpanzees in the past several million years (Hedges et al., 2004; Mills et al., 2006), human and chimpanzee Apobec3DE may have specialized to differentially restrict species-specific Alu elements. For example, humans have higher levels of Alu-Ya5 and Alu-S subfamily insertions than chimps, and both species have recent Alu-Y and -Yc1 insertions as well as remnants of ancient Alu-J elements (Bennett et al., 2008; Hedges et al., 2004; Mills et al., 2006). Therefore, I tested human and chimpanzee Apobec3DE against the consensus Alu-Y, -S, and -J subfamilies (Bennett et al., 2008). I found that human and chimpanzee Apobec3DE restricted all Alu subfamily members tested by 80-95% (Fig. 7a). I also tested the ability of human Apobec3G to inhibit Alu elements to determine whether the ability to restrict all Alu subfamilies was gained specifically by Apobec3DE. Human Apobec3G also restricted all Alu elements tested, indicating that restriction of multiple Alu subfamilies is not unique to Apobec3DE. To determine whether human Apobec3DE can non-specifically inhibit foreign DNA, as has been previously reported (Refsland et al., 2010), I co-transfected human Apobec3DE with a plasmid that expresses neomycin resistance. I did not observe any difference in the number of neomycin-resistant colonies formed in the presence or absence of human Apobec3DE (Fig. 7b). Thus, Apobec3DE specifically inhibits MusD and Alu retrotransposition. Despite the sequence divergence between human and chimpanzee Apobec3DE, both are able to restrict MusD and multiple Alu subfamilies, suggesting this is a conserved and important function of Apobec3DE that is retained despite the sequence divergence between human and chimpanzee *Apobec3DE* genes. Moreover, this emphasizes that the diversification of Alu elements was not the primary driver of the evolution of human and chimpanzee Apobec3DE.

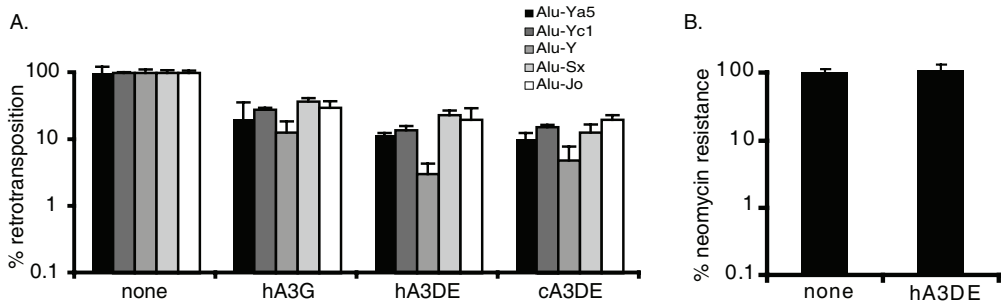


Fig. 7. Human and chimpanzee Apobec3DE restrict multiple Alu subfamilies.
Human and chimpanzee Apobec3DE were expressed in retrotransposition assays to determine their antiviral activity. Retrotransposition activity is normalized to 100% in the absence of Apobec3 (none). (A) Retrotransposition of consensus Alu subfamilies in the absence or presence of human Apobec3G (hA3G), human Apobec3DE (hA3DE), or chimpanzee Apobec3DE (hA3DE). (B) Percentage of neomycin-resistant colonies formed after transfection of retrotransposon plasmids that express neomycin resistance in the absence or presence of human Apobec3DE. Error bars represent the standard deviation of triplicate transfections within one experiment.

Chimp, but not human, Apobec3DE restricts lentiviruses

I further characterized the breadth of Apobec3DE restriction by testing the ability of human and chimpanzee orthologs to inhibit a broad range of retroviruses. I assessed their ability to inhibit MLV, HIV-1, HIV-2, and SIVagm in the presence or absence of Vif using single-round infectivity assays. I co-transfected Apobec3DE with proviral constructs and VSV-G to produce pseudotyped virions, and equivalent amounts of virions were used to infect SupT1 cells.

I found that neither human nor chimpanzee Apobec3DE restricted MLV or HIV-2Δ*vif*, unlike human Apobec3G, which restricted both MLV and HIV-2Δ*vif* (Fig. 8a and b). However, I found a marked difference between human and chimpanzee Apobec3DE in their ability to restrict other retroviruses. Chimpanzee Apobec3DE restricted HIV-1Δ*vif* and SIVΔ*vif* (20- and 40-fold, respectively, Fig. 8c and d) almost as potently as human Apobec3G, where as human Apobec3DE only weakly restricted HIV-1Δ*vif* and SIVΔ*vif* (3-fold and 6-fold, respectively). This difference between human and chimpanzee Apobec3DE in their activity against HIV-1Δ*vif*

and SIVagmΔ*vif* is significant ($p < 0.05$). Clearly, chimpanzee Apobec3DE, but not human Apobec3DE, can potently restrict certain lentiviruses.

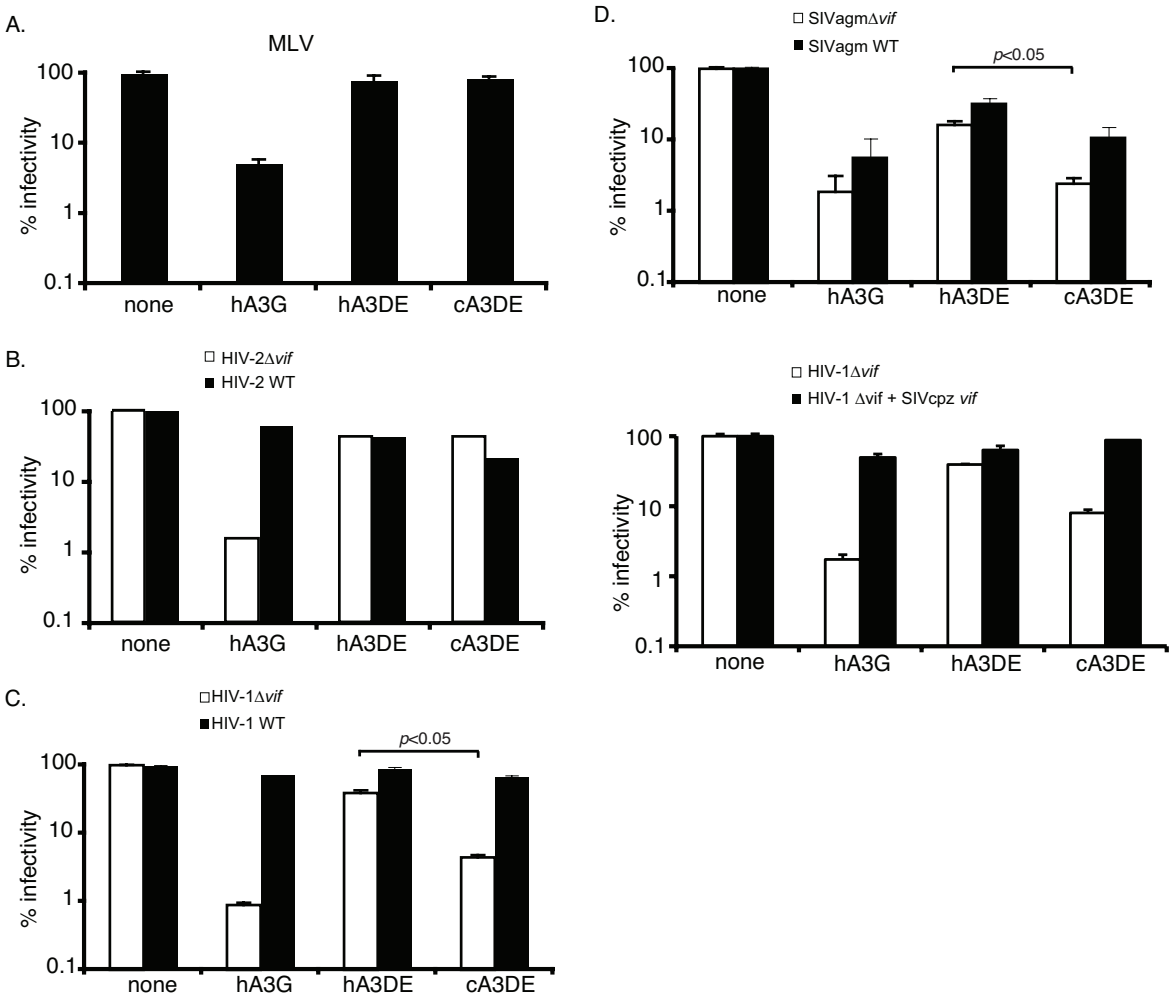


Fig. 8. Restriction of retroviruses by human and chimpanzee Apobec3DE.
Human Apobec3DE (hA3DE) and chimpanzee Apobec3DE (cA3DE) were expressed in single-round infectivity assays to determine antiviral activity. Infections in the presence of human Apobec3G (hA3G) were used as a positive control. Infectivity is normalized to 100% in the absence of Apobec3 (none). (A) Infectivity of MLV. (B) Infectivity of HIV-2 without Vif (HIV-2Δ*vif*) in open bars, and HIV-2 containing Vif (HIV-2 WT) in filled bars. (C) Infectivity of HIV-1 without Vif (HIV-1Δ*vif*) in open bars, and HIV-1 containing Vif (HIV WT) in filled bars. (D) Infectivity of SIVagmTAN without Vif (SIVagmTANΔ*vif*) in open bars, and SIVagmTAN containing Vif (SIVagmTAN WT) in filled bars. (E) Infectivity of HIV-1 without Vif (HIV-1Δ*vif*) in open bars, and HIV-1 containing Vif from SIVcpz (HIV-1Δ*vif* + SIVcpz Vif) in filled bars. Experiments were performed at least 2 times, and results from one representative experiment are shown. Error bars represent the standard deviation of triplicate infections within one experiment. *p*-values were calculated using a paired two-tailed Student's *t*-test.

Because human and chimpanzee Apobec3DE have differential antiviral activities against HIV-1Δ*vif* and SIVagmΔ*vif*, I hypothesized that they may also be differentially sensitive to neutralization by Vif proteins. However, infectivity of HIV-1 was recovered in the presence of HIV-1 and SIVcpz Vif for all Apobec3s tested (Fig. 8c and 8e), and infectivity of SIVagm was not fully recovered by any Apobec3 tested (Fig. 8d). Therefore, human and chimpanzee Apobec3DE are similar in their sensitivity to Vif proteins. These data suggest that the divergence between human and chimpanzee Apobec3DE, which allowed chimpanzee Apobec3DE to target a much broader range of retroelements than human Apobec3DE, was not driven by antagonism by Vif.

Apobec3DE is incorporated into virions and protein levels are decreased by HIV-1 Vif

I sought to determine the basis for the differences in antiviral activity between human and chimpanzee Apobec3DE. I tested a gradient of human and chimpanzee Apobec3DE levels against HIV-1 and examined, in parallel, their cellular and virion-incorporated protein levels. HIV-1Δ*vif* virions containing increasing amounts of human Apobec3DE remained infectious (Fig. 9a, filled circles, solid lines). However, HIV-1Δ*vif* virions containing increasing amounts of chimpanzee Apobec3DE became uninfectious in a dose-dependent manner (open circles, solid lines). At the highest concentration of Apobec3DE, chimpanzee Apobec3DE restricted HIV-1Δ*vif* 20 times more potently than human Apobec3DE. Western blots of cellular lysates show similar protein levels of human and chimpanzee Apobec3DE (Fig. 9b, upper panel), indicating that differential protein expression is not the cause of the differential restriction phenotypes. In addition, I examined viral lysates to determine the levels of Apobec3DE incorporated into virions and found that both human and chimpanzee Apobec3DE are incorporated into virions at

high levels (Fig. 9b, lower panel). Thus, neither limits on cellular expression nor virion incorporation underlie the differences in antiviral activity between human and chimpanzee Apobec3DE.

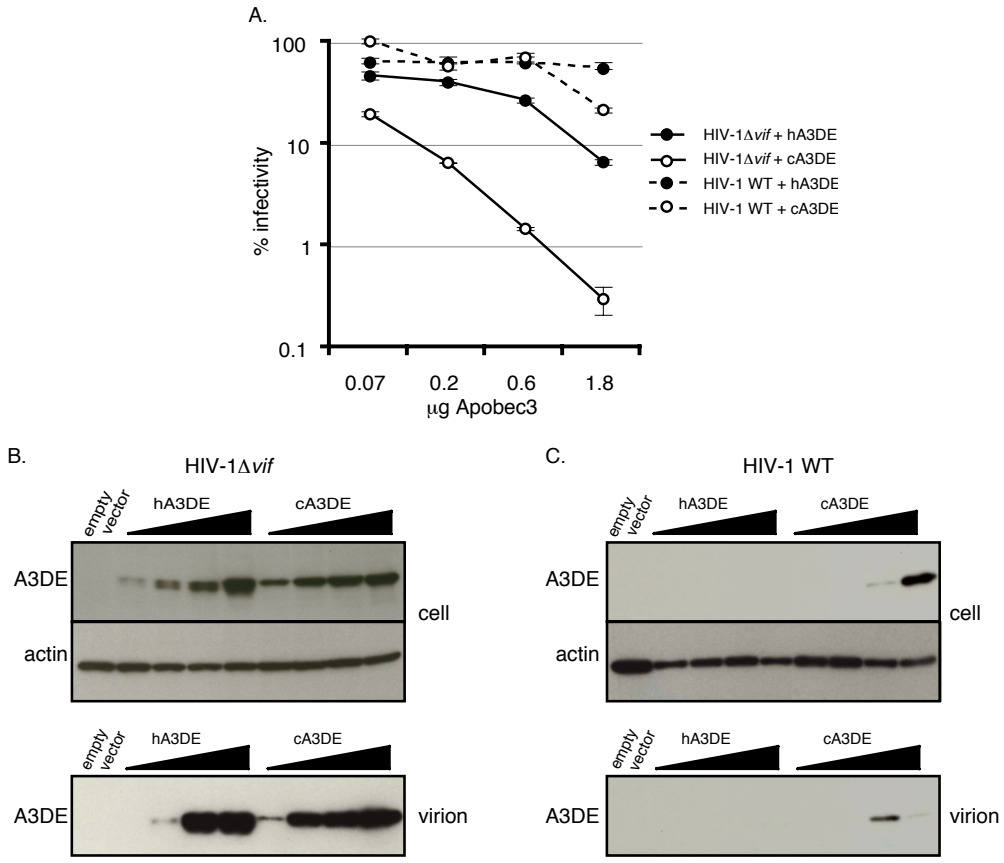


Fig. 9. Human and chimpanzee Apobec3DE are incorporated into HIV-1Δ*vif* virions.
The infectivity of HIV with and without Vif was assessed in the presence of increasing amounts of Apobec3DE. (A) Three-fold dilutions of human and chimpanzee Apobec3DE were expressed in single-round infectivity assays. Infections are represented as a percentage of the infectivity of HIV-1 without Apobec3, which was set to 100%. Filled circles are infections in the presence of human Apobec3DE (hA3DE), and open circles are infections in the presence of chimpanzee Apobec3DE (cA3DE). Solid lines are infections with HIV-1 without Vif (HIV-1Δ*vif*), and dashed lines are infections with HIV-1 containing Vif (HIV-1 WT). The experiment was performed at least 3 times, and results from one representative experiment are shown. Error bars represent the standard deviation of triplicate infections within one experiment. (B) Western blot analysis of Apobec3DE protein levels in the absence of Vif. Upper panel: Cell lysates from HIV-1Δ*vif* infections. Lower panel: Viral lysates from HIV-1Δ*vif* infections. (C) Western blot analysis of Apobec3DE protein levels in the presence of Vif. Upper panel: Cell lysates from HIV-1 infections containing Vif. Lower panel: Viral lysates from HIV-1 infections containing Vif. Apobec3DE levels are increasing from left to right on each blot. Apobec3DE was detected with an anti-HA antibody. Actin was used as a loading control for cellular lysates. Virions were normalized by quantification of p24 gag levels before lysis.

I also found that HIV-1 Vif can antagonize increasing amounts of both human and chimpanzee Apobec3DE (Fig. 9a, dashed lines). Additionally, cellular and viral levels of Apobec3DE protein drop dramatically in the presence of Vif (Fig. 9c). Consistent with previous models of Vif action (reviewed in (Malim and Emerman, 2008)), my findings suggest that HIV-1 Vif antagonizes Apobec3DE by targeting it for degradation in the cell. Thus, despite the differential ability of human and chimpanzee Apobec3DE to restrict HIV-1Δ*vif*, human and chimpanzee Apobec3DE are incorporated into virions and protein levels are decreased by HIV-1 Vif, indicating that at least some of their functional interactions with lentiviral proteins must be conserved.

The C-terminus of Apobec3DE determines anti-viral activity

To determine which region of chimpanzee Apobec3DE confers its ability to restrict lentiviruses, I made several chimeras of human and chimpanzee Apobec3DE (Fig. 10a). Swapping the last 131 amino acids, differing by 10 residues, of chimpanzee Apobec3DE with the corresponding residues of human Apobec3DE (human/chimpanzee 1) lead to a loss of the ability of chimpanzee Apobec3DE to restrict HIV-1Δ*vif* (Fig. 10b). Moreover, the reverse swap (human/chimpanzee 2) allowed human Apobec3DE to restrict HIV-1Δ*vif* ($p < 0.01$). These chimeras led to the identification of the C-terminus of the protein as responsible for its antiviral activity. Interestingly, the human/chimpanzee 2 chimera can restrict HIV-1Δ*vif* even more strongly than chimpanzee Apobec3DE ($p < 0.01$), suggesting that an interaction with the N-terminus of human Apobec3DE enhances restriction. By making additional chimeras, I found that swapping a 34 amino acid patch, differing by 5 residues, within the C-terminus of human Apobec3DE (human/chimpanzee 4) is sufficient for human Apobec3DE to gain a level of restriction of HIV-1Δ*vif* equal to chimpanzee Apobec3DE. In support of this finding, one of these residues (position

320) has recently been shown to affect the antiviral activity of several Apobec3s, including Apobec3DE (Dang et al., 2011). Therefore, this patch in the C-terminus of Apobec3DE is essential for restriction of HIV-1, and additional residues across Apobec3DE also contribute to antiviral activity. These data show that the C-terminus of Apobec3DE determines the target specificity of Apobec3DE and suggest that sequence divergence in this region is especially relevant to antiviral activity.

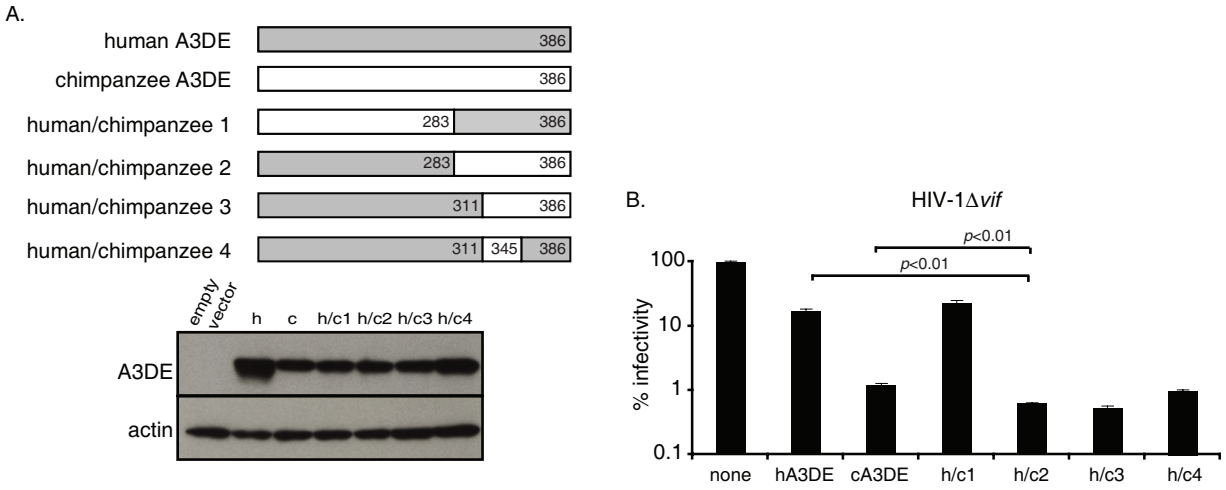


Fig. 10. C-terminus of Apobec3DE determines its ability to restrict HIV-1Δ*vif*.
(A) Upper panel: Schematic of human/chimpanzee (h/c) Apobec3DE chimeras used in infectivity assays. Lower panel: Western blot analysis of cellular levels of Apobec3DE chimeras. Apobec3DE was detected with an anti-HA antibody. Actin was used as a loading control. (B) Infectivity of HIV-1Δ*vif* in the presence of human/chimpanzee chimeras. Infections are represented as a percentage of the infectivity of HIV-1 without Apobec3 (none), which was set to 100%. Experiments were performed at least 3 times, and results from one representative experiment are shown. Error bars represent the standard deviation of triplicate infections within one experiment. *p*-values were calculated using a two-tailed Student's *t*-test.

Inefficient hypermutation of viral genomes by human Apobec3DE

Human Apobec3DE is incorporated into HIV-1Δ*vif* virions yet does not restrict viral replication (Fig. 9). Because Apobec3s can restrict viruses by inducing G-to-A hypermutation of the viral genome, I asked whether human Apobec3DE induces hypermutation of HIV-1. I infected SupT1

cells with HIV-1Δ*vif* containing human or chimpanzee Apobec3DE and measured G-to-A hypermutation of reverse-transcribed viral genomes.

I found that viral genomes had a higher rate of G-to-A hypermutation when viruses contained chimpanzee Apobec3DE compared to viruses containing human Apobec3DE (0.63% and 0.19%, respectively, Fig. 11a). Furthermore, I tested the human/chimpanzee ApobecDE chimeras from Fig. 10 for their ability to induce hypermutation of the viral genome, and the levels of hypermutation correlate well with their antiviral activity in single-round infectivity assays. For instance, human/chimpanzee 1, which has reduced levels of HIV-1 restriction, has a very low rate of G-to-A hypermutation (0.09%), where as human/chimpanzee 2 and human/chimpanzee 4, which have increased levels of restriction, induce high rates of G-to-A hypermutation (0.79% and 1.11%, respectively). As a positive control, genomes were collected from cells infected with virus containing human Apobec3G, and these sequences had very high levels of G-to-A hypermutation (1.48%), as expected. Thus, while human Apobec3DE can deaminate cytidines during infection, it induces hypermutation of the HIV-1 genome much less frequently than chimpanzee Apobec3DE does, and this difference in hypermutation is determined by the residues identified in Fig. 9 as important for antiviral activity.

The sequence context of the mutated cytidines was also examined to identify target motifs for Apobec3DE deamination. I calculated the frequency of each nucleotide at the -2, -1, +1, and +2 positions of the deaminated cytidines identified in Fig. 11a. As has been previously shown, cytidines deaminated by human Apobec3G were found most frequently with a cytidine in the -1 position, referred to as a CC motif (Fig. 11b). I found that human Apobec3DE also preferentially deaminated cytidines in a CC motif or, slightly less frequently, with a thymine in the -1 position, referred to as a TC motif. Chimpanzee Apobec3DE clearly preferentially targeted

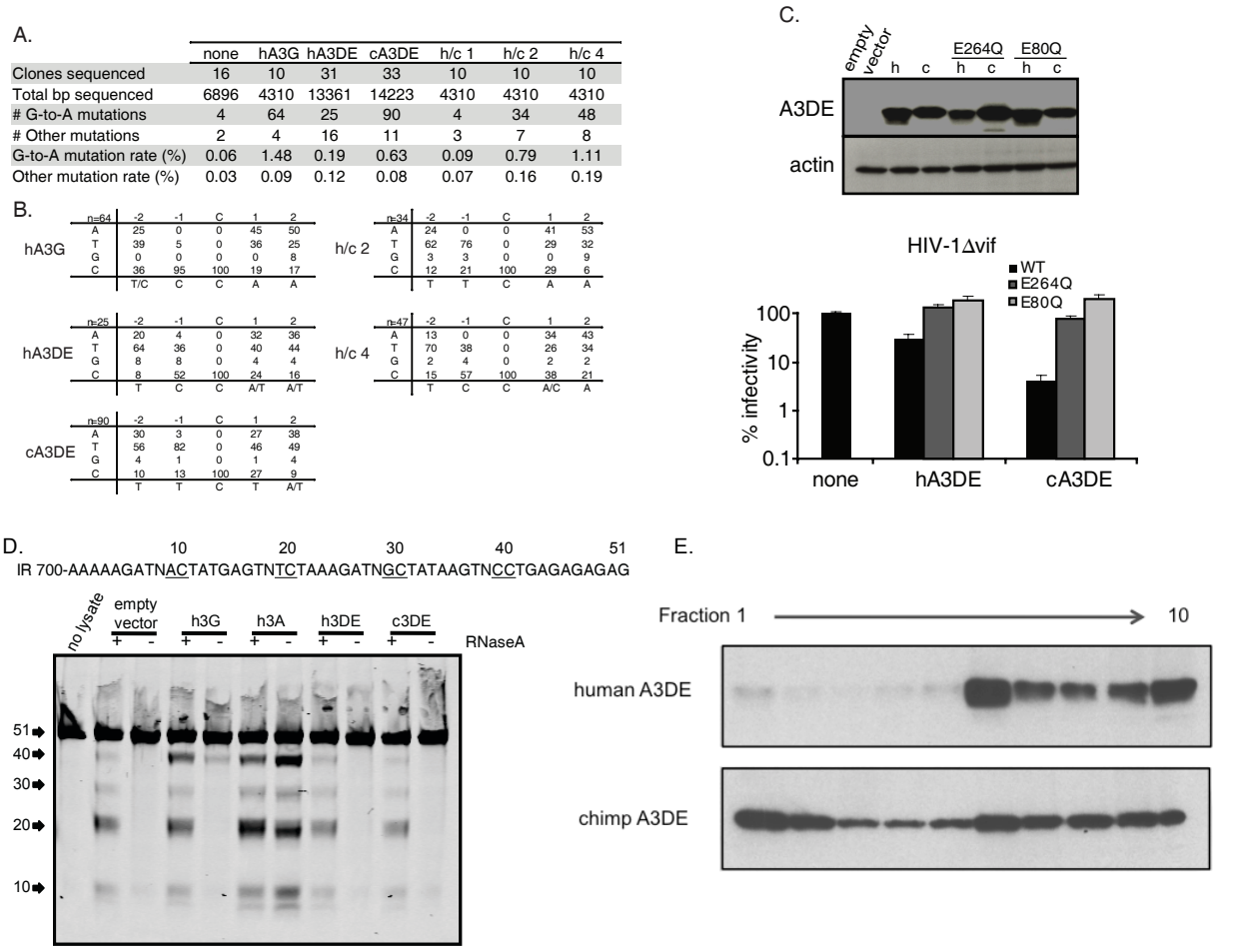
cytidines in a TC context. Human/chimpanzee chimera 2, which contains the C-terminus of chimpanzee Apobec3DE, also shared this TC target motif. Interestingly, human/chimpanzee 4, which contains only 5 residues from chimpanzee Apobec3DE, deaminated cytidines in a context more similar to human Apobec3DE (TC and CC motifs). This suggests that the target motif of Apobec3DE is determined by residues in the C-terminus outside of these 5 residues. Together, these data show that human and chimpanzee Apobec3DE target cytidines in slightly different contexts and at very different rates during infection.

To determine whether the high levels of G-to-A hypermutation induced by chimpanzee Apobec3DE contribute to its restriction of HIV-1, I evaluated the dependence of Apobec3DE restriction on its cytidine deaminase catalytic activity. I created catalytic mutants of human and chimpanzee Apobec3DE by changing a single glutamate to a glutamine within the $HxE(x)_{27}PCxxC$ catalytic domain (MacGinnitie et al., 1995; Mangeat et al., 2003). Because Apobec3DE contains two cytidine deaminase domains, I mutated each domain individually (E80Q and E264Q). I also constructed double catalytic mutants, but these mutants had very low or undetectable protein expression levels (data not shown). I found that the catalytic mutants of Apobec3DE lost the ability to restrict HIV-1Δ*vif* (Fig. 11c, shaded bars), showing that cytidine deamination is necessary for Apobec3DE restriction of HIV-1. Together, these data suggest that human Apobec3DE has weak antiviral activity against HIV-1, despite being incorporated into virions, due to lower levels of cytidine deamination of the HIV-1 genome.

To test whether human Apobec3DE has less intrinsic cytidine deaminase activity compared to chimpanzee Apobec3DE, I performed an *in vitro* gel-based deaminase activity assay. I designed an IR7000-labeled oligonucleotide that contains all possible trinucleotide cytidine motifs. Lysates from transfected 293T cells were incubated with the oligo in the

presence or absence of RNase A to allow deamination of deoxycytidine residues. RNase A was added to break up Apobec3-containing cellular ribonucleoprotein complexs that can inhibit *in vitro* enzymatic activity of Apobec3G (Chiu et al., 2005). Uracil DNA gylocosylase (UDG) was added exogenously to convert deaminated residues to abasic sites, which were then cleaved into fragments when incubated at a basic pH, and resolved by electophoresis. By quantifying the cleaved vs. uncleaved bands, I was able to determine the efficiency of Apobec3 deamination. These assays were performed with lysates from 293T cells, which appear to contain endogenous cytidine deaminases with TC, GC, and AC motifs (Fig. 11d, empty vector). As a control, I found that Apobec3G preferentially deaminated cytidines in a CC motif, and Apobec3A preferentially deaminated cytidines in a TC or CC motif. However, I was unable to detect Apobec3DE deaminase activity with this assay.

Another possibility is that human Apobec3DE has strong cytidine deaminase activity but is prevented from efficiently accessing the viral genome by cellular host factors. To test whether human and chimpanzee bind to different cellular factors, I performed velocity sedimentation of transfected Apobec3DE. I found that human and chimpanzee Apobec3DE exist in cellular complexes with different molecular masses (Fig. 11e). Thus, human and chimpanzee Apobec3DE likely bind different cellular components, which may also influence their ability to restrict lentiviruses.



	none	hA3G	hA3DE	cA3DE	h/c 1	h/c 2	h/c 4
Clones sequenced	16	10	31	33	10	10	10
Total bp sequenced	6896	4310	13361	14223	4310	4310	4310
# G-to-A mutations	4	64	25	90	4	34	48
# Other mutations	2	4	16	11	3	7	8
G-to-A mutation rate (%)	0.06	1.48	0.19	0.63	0.09	0.79	1.11
Other mutation rate (%)	0.03	0.09	0.12	0.08	0.07	0.16	0.19

Fig. 11. Chimpanzee Apobec3DE induces higher levels of hypermutation than human Apobec3DE does during viral infection.

(A). HIV-1 genomes were sequenced from HIV-1Δ*vif* infections performed in the absence of Apobec3 (none) or in the presence of human Apobec3G (hA3G), human Apobec3DE (hA3DE), chimpanzee Apobec3DE (cA3DE), or human/chimpanzee chimeras, and the G-to-A mutation rate of viral genomes was calculated. The rate of other mutations was also calculated. Combined data from three independent experiments are shown. (B) The nucleotide context of G-to-A mutations was characterized. The percentage of each nucleotide occurring at the -2, -1, +1, and +2 positions of deaminated cytidines is shown, with the most common nucleotide listed below each position. The number of mutations examined is noted in the top left corner of each grid. (C) Upper panel: Western blot analysis of cellular levels of human (h) and chimpanzee (c) Apobec3DE catalytic mutants. Apobec3DE was detected with an anti-HA antibody. Actin was used as a loading control. Lower panel: Infectivity of HIV-1Δ*vif* in the presence of wild-type Apobec3DE (WT) or Apobec3DE catalytic mutants (E264Q and E80Q). Infections are represented as a percentage of the infectivity of HIV-1 without Apobec3 (none), which was set to 100%. (D) *in vitro* deaminase assay of Apobec3 proteins' activity. The sequence for the labeled oligonucleotide is shown with dinucleotide targets underlined, and the size of cleaved fragments is written above each cytidine target. Reactions were performed with or without RNase A and resolved by gel electrophoresis. (E) Velocity sedimentation of human and chimpanzee Apobec3DE. Western blot analysis of fractions taken from the top of the gradient (fraction 1) to the bottom of the gradient (fraction 10). Apobec3DE was detected with an anti-HA antibody. Experiments were performed at least 2 times, and results from one representative experiment are shown. Error bars represent the standard deviation of triplicate infections within one experiment.

Rapid evolution of Apobec3DE in chimpanzees

Due to the differential antiviral activity found here for human and chimpanzee Apobec3DE, I was interested in whether this functional divergence was driven by positive selection. In previous evolutionary analyses of *Apobec3* genes, pairwise comparisons of human and chimpanzee divergence found a high dN/dS ratio for *Apobec3DE* (Sawyer et al., 2004). Since pairwise comparisons do not indicate which species is rapidly evolving, I performed a more detailed evolutionary analysis of *Apobec3DE* to determine whether human or chimpanzee Apobec3DE experienced a selective pressure.

I collected *Apobec3DE* sequences from additional hominoids (bonobo, gorilla, gibbon) and Old World Monkeys (rhesus macaque, African green monkey, Patas monkey). I amplified most Apobec3DE sequences from genomic DNA. However, BAC sequences containing the Apobec3 locus from gorilla (accession number FP245429), gibbon (ADFV01024628,) and orangutan (AC206461) genomes appeared to be missing the 3' half or all of *Apobec3DE*. Since I was unsure whether this represented true deletions of sequences or assembly errors, I designed primers in the 5' portion of the gene and used 3' RACE to amplify gorilla and gibbon *Apobec3DE*. I was unable to amplify orangutan *Apobec3DE* in this way as no BAC clones containing Apobec3DE sequences were identified to design effective primers. Because of the large amount of homology between *Apobec3DE* and *Apobec3F*, I also collected sequences for *Apobec3F* to distinguish the paralogs. Nucleotide sequences for *Apobec3DE* and *Apobec3F* were aligned, and a maximum likelihood (ML) phylogeny was constructed (Fig. 12a). The sequences I obtained for *Apobec3DE* (blue) and *Apobec3F* (green) cluster with known sequences and are clearly distinct. The phylogenies of *Apobec3DE* and *Apobec3F* follow the predicted species tree with high statistical support, except for the gorilla-chimpanzee-human trichotomy, which could

not be resolved with these sequences. I evaluated an alignment of *Apobec3DE* sequences for evidence of recombination using the GARD program (Kosakovsky Pond et al., 2006) and found no evidence for recombination in *Apobec3DE*.

To determine the rate of evolution of *Apobec3DE* along each primate lineage, I used phylogenetic analysis by maximum likelihood (PAML) and performed a free-ratio model to calculate the ratio of the nonsynonymous mutation rate (dN) to the synonymous mutation rate (dS) for each branch (called dN/dS). Because the gorilla-chimpanzee-human trichotomy was ambiguous for Apobec3DE, I performed these analyses using two phylogenies of *Apobec3DE*—the gene phylogeny obtained by ML with gorilla *ApobecDE* clustering within human *Apobec3DE* (Fig. 12b, right) and a species phylogeny that clusters human *Apobec3DE* within gorilla *Apobe3DE* (Fig. 12b, left). For both input trees, I found that a model of codon evolution under positive selection (M2 and M8) fit our data more strongly than a model of codon evolution under neutral selection (M1 and M7, $p < 0.001$, Table 3). Also, the dN/dS values along each lineage are similar between the two trees. I found that the chimpanzee-bonobo ancestor of *Apobec3DE* has a dN/dS value much greater than 1 (6.6 or 6.5), indicative of rapid evolution in this lineage between 2 and 6 million years ago. This represents the most intense episode of positive selection among all primate Apobec3DEs and, indeed, among all Apobec3s (OhAinle et al., 2006; Sawyer et al., 2004). The human *Apobec3DE* lineage exhibits lower dN/dS values (0.9 or 2.3). More than twice as many amino acids changed in chimpanzee *Apobec3DE* compared to human *Apobec3DE* since their divergence (25 or 26 compared to 11 or 10, respectively). To assess the dN/dS ratio of the chimpanzee-bonobo ancestor of *Apobec3DE*, I compared a model where dN/dS was fixed at 1 along this branch to a model where dN/dS was allowed to be greater than 1 along this branch. The second model fit my data more strongly ($p < 0.05$), confirming

positive selection along this lineage. The 25 sites that unambiguously changed in the chimpanzee lineage were identified using the CODEML marginal reconstruction of the human-chimpanzee ancestral sequences from both phylogenies (residues highlighted in Fig. 12c). Strikingly, all of the residues identified in Fig. 10 that alter Apobec3DE antiviral activity evolved in the chimpanzee lineage, rather than in the human lineage. This shows that the increase in antiviral activity within chimpanzee Apobec3DE was driven by positive selection since separation from the human-chimpanzee common ancestor.

I also identified individual codons in Apobec3DE under positive selection across primates. I found that 15 residues have a high posterior probability of dN/dS > 1 using both phylogenies (*BEB > 0.95, **BEB > 0.99, Fig. 12c). Interestingly, many of the positively-selected codons cluster in the C-terminus of Apobec3DE, which contains residues important for viral restriction. My analysis of the evolution of *Apobec3DE* indicates that selection has acted both across primates and very potently in the chimpanzee lineage. Further, this evolution is targeted to the domain of Apobec3DE that determines viral specificity, suggesting that positive selection may have driven changes in the antiviral activity of other primates' Apobec3DE in addition to broadening the target range of chimpanzee Apobec3DE.

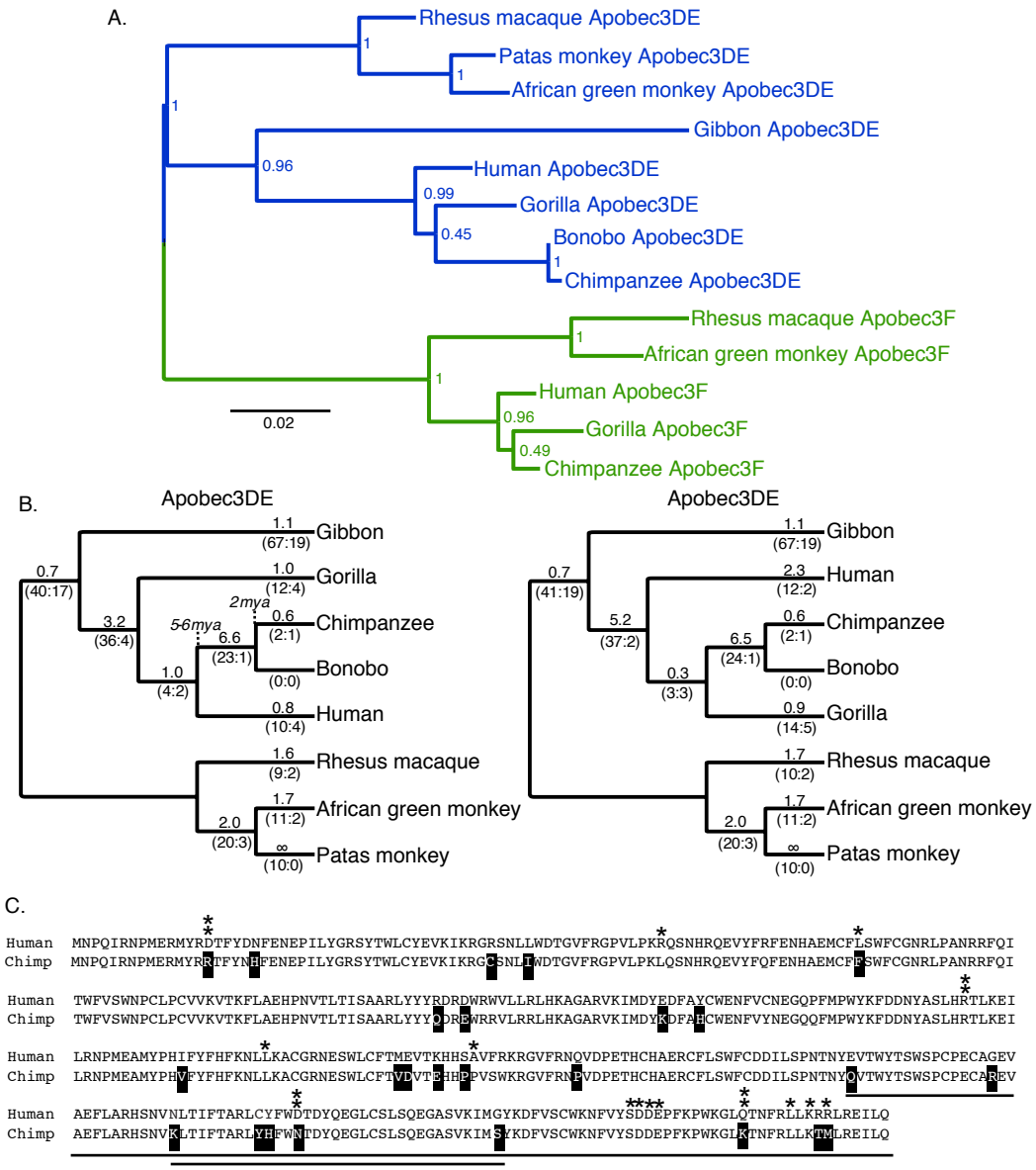


Fig. 12. *Apobec3DE* evolved rapidly in chimpanzee ancestors.
Sequences obtained by PCR in this study, as described in Materials and Methods, were analyzed. (A) A maximum-likelihood phylogeny based on primate *Apobec3DE* and *Apobec3F* sequences is shown. Statistical support, calculated as aLRT values, is shown for each node. (B) Positive selection analysis of Apobec3DE. Two phylogenies of *Apobec3DE*, one with the human orthologue as an outgroup to gorilla, chimpanzee, and bonobo *Apobec3DE* (right) and another with gorilla *Apobec3DE* as an outgroup to human, chimpanzee, and bonobo *Apobec3DE* (left), were analyzed. Global dN/dS ratios are indicated above each branch. The number of non-synonymous and synonymous changes along each branch are indicated below each branch in parentheses (NS:S). dN/dS values greater than 1 are indicative of positive selection. (C) Codons in Apobec3DE under positive selection. An alignment of human and chimpanzee Apobec3DE protein sequences is shown with chimpanzee-specific changes highlighted in black. Residues having a high probability of being under positive selection across all primates are marked with asterisks. Residues with BEB values >0.95 are marked with a single asterisk, and residues with BEB values >0.99 are marked with 2 asterisks. The C-terminal domain in Apobec3DE that was tested for antiviral activity is underlined, with the minimal 5 residues that are required for antiviral activity underlined twice.

Codon model[a]	M1 vs M2[b]	M7 vs M8[c]	chimpanzee-bonobo ancestor[d]
(A) (h, (g, (c, b)))			
F3x4	$p<0.001$	$p<0.001$	$p<0.05$
F61	$p<0.001$	$p<0.001$	$p<0.05$
(B) (g, (h, (c, b)))			
F3x4	$p<0.001$	$p<0.001$	$p<0.05$
F61	$p<0.001$	$p<0.001$	$p<0.05$

Table 3. Likelihood ratio tests for positive selection in *Apobec3DE*
(A) A tree obtained with *Apobec3DE* gene sequences, in which human *Apobec3DE* (h) is an outgroup to gorilla (g), chimpanzee (c), and bonobo (b) *Apobec3DE*, was used for analyses.
(B) A tree following species phylogeny, in which gorilla *Apobec3DE* is an outgroup to human, chimpanzee, and bonobo *Apobec3DE*, was used for analyses.
[a] Two different models of codon frequencies (F3x4 and F61) were used for analyses.
[b] Likelihood ratio tests were performed by comparing model 1 (two-state, dN/dS > 1 disallowed) to model 2 (two-state, dN/dS > 1 allowed). *p* values < 0.05 indicate that a model of selection (M2) provides a better fit of the data.
[c] Likelihood ratio tests were performed by comparing model 7 (beta distribution, dN/dS > 1 disallowed) to model 8 (beta distribution, dN/dS > 1 allowed). *p* values < 0.05 indicate that a model of selection (M8) provides a better fit of the data.
[d] Likelihood ratio tests were performed by comparing a model where dN/dS = 1 to a model where dN/dS > 1 for the chimpanzee-bonobo ancestor. *p* values < 0.05 indicate that a model of selection (dN/dS > 1) provides a better fit of the data.

Discussion

In this study, I investigated the role and diversification of the primate gene *Apobec3DE* in host defense. I found that the human and chimpanzee orthologs of Apobec3DE are able to strongly inhibit Alu and MusD retrotransposition. Chimpanzee Apobec3DE is also able to restrict the infectivity of HIV-1 and SIVagmTAN in the absence of Vif, while human Apobec3DE has weak activity against all retroviruses tested. I show that human and chimpanzee Apobec3DE have similar cellular and virion-associated protein levels but differentially hypermutate HIV-1 genomes. Further, I show that *Apobec3DE* has evolved very rapidly in chimpanzee ancestors and in the domain responsible for antiviral activity. This suggests that a selective pressure existing

between 2 and 6 million years ago drove the adaptation of antiviral activity of chimpanzee Apobec3DE.

The C-terminus of Apobec3DE contains the residues responsible for the differential activity of human and chimpanzee Apobec3DE (Fig. 10). I mapped this region to a patch between residues 311 and 345 of Apobec3DE containing five amino acid differences. This data correlates very well with recently published data showing that human Apobec3DE, which has weak activity against HIV-1, has much higher antiviral activity when the cysteine residue at position 320 is changed to a tyrosine (notably, the residue found in chimpanzee Apobec3DE) (Dang et al., 2011). Additionally, many residues in the C-terminus of Apobec3DE are under positive selection (Fig. 12), suggesting that this domain of Apobec3DE has repeatedly changed throughout evolutionary history, perhaps because it forms an interaction with a viral factor.

The reason that human Apobec3DE does not induce hypermutation of lentiviruses is unclear, as the human Apobec3DE protein is efficiently packaged into virions (Fig. 9) and can restrict other retroelements (Fig. 6). Potentially, the human Apobec3DE protein may have low levels of enzymatic deaminase activity compared to chimpanzee Apobec3DE; however, I was unable to address this because, in the *in vitro* gel-based deaminase assay, I was unable to detect deaminase activity of either human or chimpanzee Apobec3DE (Fig. 11). Alternatively, human Apobec3DE may contain catalytic activity but be improperly targeted within the virion, such as is proposed for human Apobec3A (Aguiar et al., 2008; Goila-Gaur et al., 2007), rendering it unable to access the viral genome appropriately. Because of the strong denaturants required to resolve human Apobec3DE by SDS-PAGE (Fig. 5) and the differences in velocity sedimentation between human and chimpanzee Apobec3DE (Fig. 11), the human Apobec3DE protein likely

has some biochemical differences from chimpanzee Apobec3DE, perhaps preventing it from binding necessary viral factors.

It is possible that human Apobec3DE lost antiviral activity since divergence with chimpanzees due to a lack of selective pressure; however, human Apobec3DE can inhibit Alu elements that have been active in human ancestors for millions of years (Fig. 6 and 7, reviewed in (Batzer and Deininger, 2002)), and this interaction likely exerted some selective pressure on Apobec3DE. The most likely scenario is that chimpanzee Apobec3DE gained antiviral activity since its divergence from human Apobec3DE. There are no changes in the C-terminus of Apobec3DE in the human lineage, suggesting that the human-chimpanzee ancestor of Apobec3DE would function most similarly to human Apobec3DE and would not have the ability to restrict lentiviruses. Because chimpanzee Apobec3DE has evolved antiviral activity only against some lentiviruses rather than against other retroviruses or additional retrotransposons (Fig. 6 – 8), the selection on chimpanzee Apobec3DE would most likely have been an ancient lentivirus. A Vif-like protein is also unlikely to have been the viral factor driving selection on chimpanzee Apobec3DE because both human and chimpanzee Apobec3DE are sensitive to antagonism by lentiviral Vif proteins. The identity of this paleolentivirus is unknown because modern primate lentiviruses are too young to have driven the rapid evolution of chimpanzee Apobec3DE, which we showed in Fig. 6 occurred 2-6 mya (Emerman and Malik, 2010; Sawyer et al., 2004). However, since endogenous lentiviral sequences identified in lemur genomes show that lentiviruses have circulated in primates for millions of years (Gifford et al., 2008; Gilbert et al., 2009), it is likely that ancient lentiviruses also infected other primates millions of years ago. Since chimpanzee Apobec3DE can restrict modern-day lentiviruses, we can infer that an extinct lentivirus probably infected chimpanzee ancestors 2-6 mya and was responsible for the selection

of Apobec3DE along the chimpanzee lineage since its divergence from the human-chimpanzee common ancestor. Thus, I have identified the possible consequences of an ancient lentivirus infection of chimpanzees, as well as a potential role for Apobec3DE in host defense in modern hominids.

Chapter IV
Loss-of-function polymorphisms in the Apobec3 locus in human populations

Introduction

Many *Apobec3* genes, including *Apobec3DE, Apobec3G,* and *Apobec3H*, have evolved under positive selection for millions of years in primates (Duggal et al., 2011; OhAinle et al., 2006; Sawyer et al., 2004). At least one *Apobec3* gene within an Old World monkey species has acquired population-specific polymorphisms in recent history that allow host evasion from lentiviral infection (Compton et al., 2012), suggesting that *Apobec3* genes may have rapidly evolved in recent primate history as well.

Within humans, several *Apobec3* genes are known to have common polymorphisms that render them defective. For example, two variants of *Apobec3H* encode proteins with decreased expression (OhAinle et al., 2008), and deletions of *Apobec3B* are common in some human populations (Kidd et al., 2007). However, whether variants that have functional consequences are present in other *Apobec3* genes is unknown. Moreover, the only allele of the human *Apobe3DE* gene that has been functionally tested for antiviral activity is less active against lentiviruses than the orthologous chimpanzee *Apobec3DE* gene (Duggal et al., 2011). Because of widespread interest in understanding the contributions of human genetics to viral susceptibility and disease, I sought to comprehensively identify and functionally characterize variants of *Apobec3* genes in human populations.

In this chapter, I identify 21 amino acid-altering mutations in the *Apobec3* locus using the 1000 Genome Project dataset, of which 9 are novel. I find that 6 common (minor allele frequency, MAF >1%) single nucleotide variants (SNVs) in Apobec3A, C, F, and G have no effect on antiviral activity. However, I find that two novel SNVs in Apobec3DE decrease

antiviral activity against HIV-1 and Alu retrotransposons. I also confirm two rare (MAF <1%) mutations in Apobec3G and Apobec3F that drastically lower endogenous protein expression. To better understand the evolutionary pressures acting on the *Apobec3* locus, I perform neutrality test statistics and find that *Apobec3DE* is evolving under purifying selection in humans. These results highlight the important role of Apobec3DE in host defense and suggest that Apobec3DE variants may be only slightly deleterious and maintained at low frequencies.

Results

Common polymorphisms in Apobec3A, C, F, and G do not affect antiviral activity.

We accessed the 1000 Genome Project Phase I genotypes for single nucleotide variants in the coding regions of each *Apobec3* gene. This dataset includes 913 geographically diverse individuals, which can be broadly grouped into three populations: African, Asian, and European ancestries. We also obtained insertions and deletions for 911 of the same individuals from the 1000 Genome Project Integrated Phase 1 release. A summary of the most common (MAF > 1%) codon-altering variants found in *Apobec3A* through *H* is found in Table 4, with the combined and population-specific frequencies of the derived allele reported.

Apobec3	Variant	ID	African (n[a] = 246)	Asian (n = 286)	European (n = 381)	Combined (n = 913)	Citation
A	T19A	rs17000556	10.6	0	0.1	2.9	
B	K62E	rs2076109	68.1	63.6	62.2	64.2	
B	P98L	rs59708943	16.3	0	5.4	6.6	
B	T146K	rs5995649	5.9	2.1	3.9	3.9	
B	L189P	rs146055882	10.6	0	0	2.8	
B	A254V	rs138093253	5.7	0	0.9	1.9	
B	CNV[b]	CNV_36030	6.1	31.1	7.8[c]	14.6[d]	(Kidd et al., 2007)
C	S188I	rs112120857	9.6	0	0	2.6	
DE	R97C	rs75858538	4.7	0	0	1.3	
DE	R248K	rs61748819	11.6	0	0	3.1	
F	A108S	rs2020390	30.1	70.8	47.2	50.0	(Mulder et al., 2010)
F	V231I	rs2076101	20.1	70.8	47.2	47.3	(Mulder et al., 2010)
F	Y307C	rs12157816	3.7	0	1.1	1.4	(Mulder et al., 2010)
G	H186R	rs8177832	37.4	7.9	3.3	13.9	(An et al., 2004)
G	E275Q	rs17496046	14.6	1.9	4.3	6.4	(Reddy et al., 2010)
H	R18L	rs139293	6.1	16.1	31.1	19.7	(OhAinle et al., 2008)
H	R105G	rs139297	15.9	67.7	52.2	47.3	(OhAinle et al., 2008)
H	E121K	rs139298	13.8	68.0	51.8	46.7	(OhAinle et al., 2008)
H	E121D	rs139299	86.8	32.3	48.4	53.7	(OhAinle et al., 2008)
H	E178D	rs139302	15.4	67.5	52.2	47.1	(OhAinle et al., 2008)
H	Del[e] 15	rs59165009	33.1	24.8	35.2[c]	31.3[d]	(OhAinle et al., 2008)

Table 4. Derived allele frequency of common variants in *Apobec3* genes.
[a] n = number of individuals. [b] CNV = copy number variant. [c] n = 379. [d] n = 911. [e] Del = deletion.

The 1000 Genome Project dataset contains 11 previously reported variants in the Apobec3 locus, including the deletion of Apobec3B (Kidd et al., 2007), the deletion of residue 15 in Apobec3H (OhAinle et al., 2008), and SNVs in Apobec3F (A108S, V231I, Y307C) (Mulder et al., 2010), Apobec3G (H186R, Q275E) (An et al., 2004; Reddy et al., 2010), and Apobec3H (R105G, E121K/D, D178E) (OhAinle et al., 2008). In addition, 9 novel variants can also be found in this dataset, including SNVs in Apobec3A (T19A), Apobec3B (K62E, P98L, T146K, L189P, A254V), Apobec3C (S188I), and Apobec3DE (R97C, R248K). Consistent with previous reports of variants from HapMap populations, polymorphisms at positions 108 and 231 in Apobec3F and at positions 105, 121, and 178 in Apobec3H are found at very high frequencies

outside of Africa. In contrast, SNVs in Apobec3A, Apobec3B, Apobec3C, Apobec3DE, Apobec3G are found at their highest frequency in African populations. The deletion of Apobec3B is most common in Asian populations, and the deletion of residue 15 in Apobec3H is found at similar frequencies in all populations, as previously reported for other datasets. All common variants in Apobec3H have been well described functionally in previous studies: the deletion of residue 15 decreases protein stability; the R105G mutation decreases protein stability and alters cellular localization; the E121K mutation decreases sensitivity to HIV-1 Vif; and the D178E mutation has no effect on protein function (Harari et al., 2009; Li et al., 2010; OhAinle et al., 2008). The loss of Apobec3B in many individuals suggests that Apobec3B is not important for human health, and the insensitivity of Apobec3B to HIV-1 Vif suggests it is not relevant to HIV-1 infection. Therefore, for the rest of my functional analyses of SNVs in Apobec3 genes, I focused on those found in Apobec3A, C, DE, F, and G.

Human Apobec3A and Apobec3C potently inhibit the replication of human retrotransposons (Bogerd et al., 2006b; Muckenfuss et al., 2006). Apobec3A T19A and Apobec3C S1881 mutations are present at frequencies near 10% in African populations and <1% in other populations (Table 4). As no SNVs in either paralog have been previously described, I tested whether these Apobec3A or Apobec3C variants affect antiviral activity. Each of these mutations was introduced into a plasmid containing epitope-tagged Apobec3A or Apobec3C. I co-transfected increasing amounts of Apobec3 plasmids with a plasmid containing the non-LTR human retrotransposon LINE-1, which expresses neomycin resistance after retrotransposition. After selection in G418, neomycin-resistant colonies were counted as a measure of LINE-1 activity. As seen by Western blot, Apobec3A T19A and Apobec3C S188I had slightly decreased protein expression compared to wild-type Apobec3A or Apobec3C protein (Fig. 13A and 13B,

upper panels). However, the mutations did not affect the ability of Apobec3A or Apobec3C to restrict the activity of LINE-1 in a dose-dependent manner (Fig. 13A and 13B, lower panels). These data suggest that the novel SNVs in Apobec3A and Apobec3C do not alter their antiviral capacity.

Human Apobec3F has potent antiviral activity against HIV-1Δ*vif* and is sensitive to HIV-1 Vif (Bishop et al., 2004; Liddament et al., 2004; Wiegand et al., 2004; Zheng et al., 2004), but reports vary in its ability to inhibit LINE-1 (Bogerd et al., 2006b; Hulme et al., 2007; Stenglein and Harris, 2006). In this dataset, Apobec3F Y307C is present at a low frequency in African and European populations (MAF <5%) and is absent in Asian populations. Previous reports have shown that Apobec3F Y307C has reduced antiviral activity (Dang et al., 2011) and increased sensitivity to HIV-1 Vif (Mulder et al., 2010). In addition, Apobec3F A108S and V231I are present at intermediate to high frequencies (MAF 20 – 70%) in all human populations, with the highest frequencies found in individuals outside of Africa. We hypothesized that these common variants of Apobec3F may explain the discrepancy in reported anti-LINE activity. Therefore, I tested these common SNVs in Apobec3F for antiviral activity. As these two mutations are in strong linkage disequilibrium (LD) in all populations ($r^2 = 0.91$), I introduced the mutations into the same Apobec3F plasmid and functionally assayed them together. Plasmids containing Apobec3F were co-transfected at multiple concentrations with HIV-1 proviral constructs containing a luciferase gene and VSV-G to produce pseudotyped virions. Equivalent amounts of virions were used to infect SupT1 cells, and viral infectivity was determined by luciferase activity of infected cells. While these mutations slightly increased protein expression of transfected Apobec3F (Fig. 13C, upper panel), they did not alter the ability of Apobec3F to restrict HIV-1Δ*vif* (Fig. 13C, lower panel, solid lines) or HIV-2Δ*vif* (data not shown), and they

did not alter their sensitivity to degradation by HIV-1 Vif (Fig. 13C, lower panel, dashed lines) or HIV-2 Vif (data not shown). We next asked if wild-type or mutant Apobec3F could inhibit LINE-1 retrotransposition. As seen in Fig. 3C, neither wild-type nor mutant Apobec3F restricted LINE-1 by more than 3-fold, even at the highest concentration. From these data, I conclude that Apobec3F does not inhibit LINE-1 and that the common SNVs in Apobec3F do not alter antiviral activity.

Human Apobec3G has the most potent antiviral activity against HIV-1Δ*vif*, is highly sensitive to HIV-1 Vif (Bishop et al., 2004; Zheng et al., 2004), and restricts Alu elements (Esnault et al., 2005; Hulme et al., 2007). Similarly to other groups, I find Apobec3G H186R and Q275E variants at their highest frequency in African individuals (MAF 37% and 15%, respectively), compared to lower frequencies (MAF <10% or <5%, respectively) in other populations in this dataset. Apobec3G H186R has been shown to restrict HIV-1Δ*vif* and be sensitive to HIV-1 Vif (An et al., 2004), but Apobec3G Q275E has not been described functionally. As these SNVs are in linkage equilibrium ($r^2 < 0.01$), I introduced these mutations into separate plasmids containing epitope-tagged Apobec3G. When transfected into 293T cells, Apobec3G H186R and Apobec3G Q275E had similar protein expression levels as compared to wild-type Apobec3G (Fig. 13D, upper panel). I assayed the anti-HIV activity of the H186R and Q275E mutants compared to wild-type Apobec3G and found that all Apobec3Gs restricted HIV-1Δ*vif* infectivity to similar levels in a dose-dependent manner (Fig. 13D, lower panel, solid lines). Apobec3G H186R and Apobec3G Q275E were also highly sensitive to HIV-1 Vif, as infectivity was recovered to 90-100% in the presence of Vif at all concentrations of Apobec3G (Fig. 13D, lower panel, dashed lines). These data suggest that variants of Apobec3G do not have altered antiviral function.

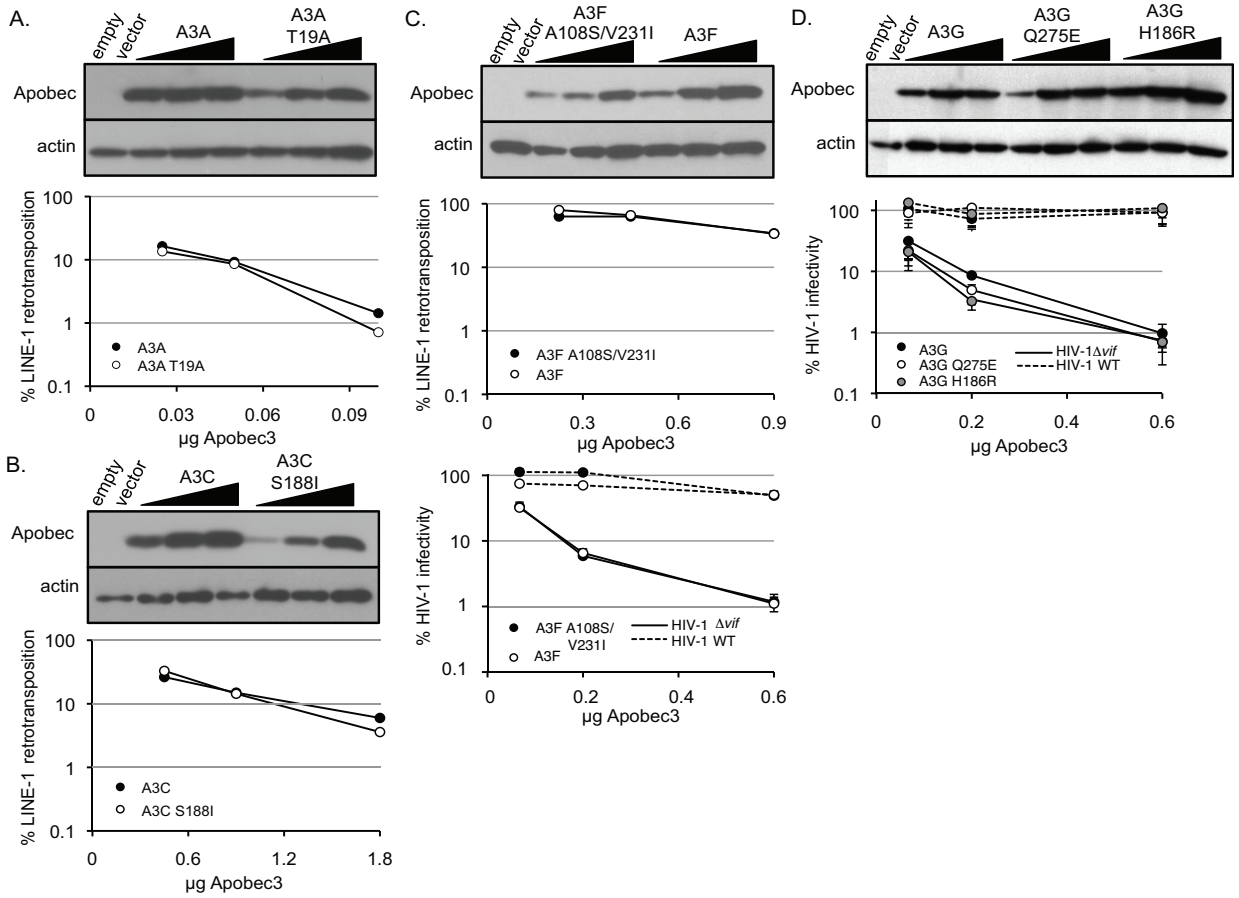


Fig. 13. Common variants in Apobec3A, C, F, and G do not affect antiviral activity.
The antiviral activity of Apobec3 mutants against HIV-1, Alu, or LINE-1 elements was compared to the antiviral activity of wild-type Apobec3. Infectivity and retrotransposition are represented as a percentage of virus infectivity or retrotransposon activity in the absence of Apobec3, which was normalized to 100%. Infections were performed with dilutions of Apobec3, and Apobec3 levels are increasing from left to right. Cellular Apobec3 protein was detected by Western blot using an anti-HA antibody, and actin was used as a loading control. (A) Apobec3A T19A. (Top) Western blot analysis of cellular Apobec3A (A3A) protein levels. (Bottom) LINE-1 retrotransposition. Filled circles are wild-type Apobec3A, and open circles are Apobec3A T19A. (B) Apobec3C S1881. (Top) Western blot analysis of cellular Apobec3C (A3C) protein levels. (Bottom) LINE-1 retrotransposition. Filled circles are wild-type Apobec3C, and open circles are Apobec3C S188I. (C) Apobec3F A108S/V231I. (Top) Western blot analysis of cellular Apobec3F (A3F) protein levels. (Middle) HIV-1 infectivity. Open circles are wild-type Apobec3F, and filled circles are Apobec3F A108S/V231I. Dashed lines are infections with HIV-1 containing Vif (HIV-1 WT), and solid lines are infections with HIV-1 lacking Vif ((HIV-1Δ*vif*). (Bottom) LINE-1 retrotransposition. (D) Apobec3G Q275E and Apobec3G H186R. (Top) Western blot analysis of cellular Apobec3G (A3G) protein levels. (Bottom) HIV-1 infectivity. Filled circles are wild-type Apobec3G, open circles are Apobec3G Q275E, and gray circles are Apobec3G H186R. Dashed lines are infections with HIV-1 containing Vif (HIV-1 WT), and solid lines are infections with HIV-1 lacking Vif (HIV-1Δ*vif*). Experiments were performed at least 3 times, and results from a representative experiment are shown.

Two polymorphisms in Apobec3DE decrease antiviral activity.

Human Apobec3DE has weak antiviral activity against HIV-1 in the absence of Vif, is highly sensitive to degradation by HIV-1 Vif, and has strong anti-retroelement activity against Alu elements (Dang et al., 2006; Duggal et al., 2011). We wanted to identify SNVs of Apobec3DE in human populations to determine whether any individuals have an Apobec3DE allele with stronger anti-HIV activity. In the 1000 Genome Project dataset, I found that the most common non-synonymous SNVs in Apobec3DE were an arginine to cysteine change at position 97 (R97C) and an arginine to lysine change at position 248 (R248K), which exist at 1.3% and 3.1% frequencies, respectively, in the combined populations. These non-synonymous SNVs are found exclusively in individuals with African ancestry at low frequencies (MAF 4.7% and 11.6%).

As these novel SNVs in Apobec3DE are in linkage equilibrium ($r^2 < 0.01$), I introduced the R97C and R248K mutations into separate plasmids containing epitope-tagged Apobec3DE. These plasmids were transfected into 293T cells at multiple concentrations, and I compared wild-type and mutant Apobec3DE for protein expression by Western blot. I found that Apobec3DE R97C has similar protein levels as wild-type Apobec3DE, while Apobec3DE R248K has drastically lower (approximately 9-fold) protein expression than wild-type Apobec3DE (Fig. 14A). I next assessed the two Apobec3DE mutants for antiviral activity in a single-round HIV-1 infectivity assay. Apobec3DE R97C and Apobec3DE R248K restricted HIV-1Δ*vif* by only 5- and 2-fold, which was significantly lower than wild-type Apobec3DE, which restricted HIV-1Δ*vif* by nearly 30-fold ($p<0.05$, Fig. 14C, filled bars). The decreased antiviral activity of Apobec3DE R97C and Apobec3DE R248K was observed at all Apobec3 concentrations tested (Fig. 14B). Similar to wild-type Apobec3DE, Apobec3DE R97C and Apobec3DE R248K were highly sensitive to HIV-1 Vif, as infectivity was rescued to 100% in the presence of HIV-1 Vif,

and were partially sensitive to HIV-2 Vif, as infectivity was rescued to 20 – 40% in the presence of HIV-2 Vif (Fig. 14C, open and gray bars). Apobec3G was used as a control for Vif sensitivity, as infectivity was rescued to nearly 100% by Vif from HIV-1 and HIV-2. Thus, novel variants of Apobec3DE have significantly decreased antiviral activity against HIV-1 but similar sensitivities to Vif.

While the R97C and R248K variants of Apobec3DE are independent mutations, they have the same functional effect of reducing the anti-HIV potential of Apobec3DE. By determining the combined frequency of alleles containing either the R97C or the R248K mutation, I find that 3.4% of Apobec3DE alleles in humans are deleterious with regards to blocking HIV replication. Within individuals of African ancestry, 16.2% of Apobec3DE alleles are deleterious (Fig. 14D), and 4.5% of individuals with African ancestry have two deleterious Apobec3DE alleles. Previous reports have shown that Apobec3DE mRNA is expressed in $CD4^+$ cells and macrophages and that it is moderately induced by interferon treatment (Koning et al., 2009; Refsland et al., 2010). Furthermore, a recent report has shown that, by knocking down Apobec3DE mRNA expression in a T cell line, HIV-1Δ*vif* replication is increased and viral hypermutations typically induced by Apobec3s are decreased (Refsland et al., 2012). It is therefore possible that the polymorphisms in Apobec3DE could impact viral replication *in vivo*.

I did not find any common (MAF >1%) polymorphisms in human Apobec3DE with increased antiviral activity relative to the wild-type allele. Therefore, we conclude that the previously reported reduced antiviral activity of human Apobec3DE relative to chimpanzee Apobec3DE (Duggal et al., 2011) is not a reflection of ascertainment bias. In fact, common variants of Apobec3DE in human populations are less effective at restricting HIV-1 than wild-type Apobec3DE.

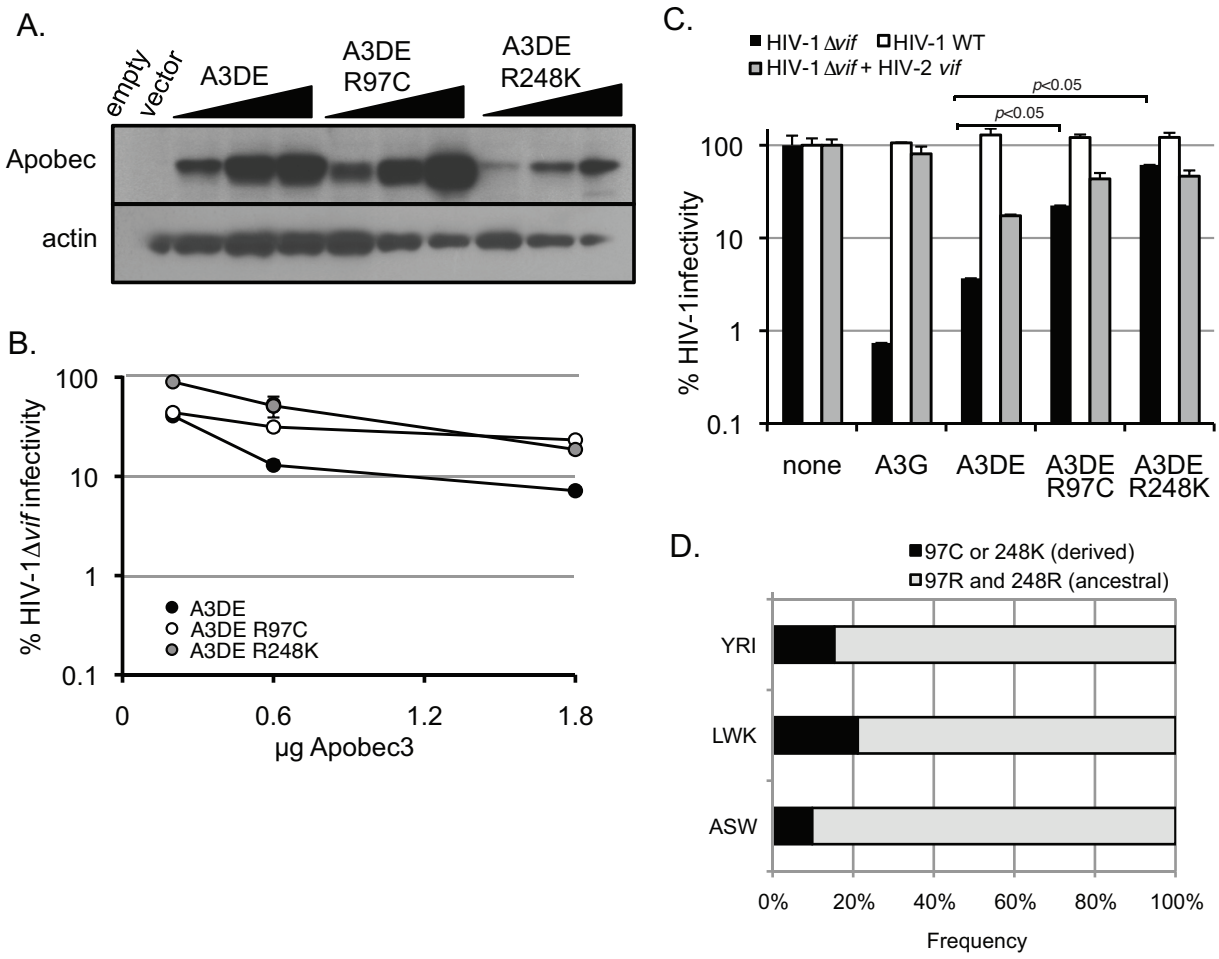


Fig. 14. Single nucleotide variants in Apobec3DE decrease antiviral activity.
The antiviral activity of Apobec3DE R97C and R248K mutants against HIV-1 was compared to the antiviral activity of wild-type Apobec3DE. Infectivity is represented as a percentage of virus infectivity in the absence of Apobec3DE (A3DE), which was normalized to 100%. (A) Western blot analysis of cellular Apobec3DE protein levels. Apobec3DE was detected using an anti-HA antibody, and actin was used as a loading control. Transfections were performed with 3-fold dilutions of Apobec3DE, and Apobec3DE levels are increasing from left to right. (B) HIV-1 infectivity. Infections were performed with HIV-1 lacking Vif (HIV-1Δ*vif*). Wild-type Apobec3DE is represented by filled circles, Apobec3DE R97C is represented by open circles, and Apobec3DE R248K is represented by gray circles. Infections were performed with 3-fold dilutions of Apobec3DE, and Apobec3DE levels are increasing from left to right. (C) HIV-1 infectivity. Open bars are infections with HIV-1 lacking Vif (HIV-1Δ*vif*), filled bars are infections with HIV-1 containing Vif (HIV-1 WT), and gray bars are infections with HIV-1 lacking Vif but containing HIV-2 Vif (HIV-1Δ*vif* + HIV-2 Vif). (D) Distribution of deleterious Apobec3DE alleles in African populations. The proportion of deleterious Apobec3DE alleles (97C or 248K) is shown in filled bars, and the proportion of wild-type Apobec3DE alleles (97R and 248R) is shown in gray bars. YRI, Yoruba individuals from Ibadan, Nigeria; LWK, Luhya individuals in Webuye, Kenya; ASW, African ancestry individuals from southwest US. Experiments were performed at least 3 times, and results from a representative experiment are shown.

Rare polymorphisms that are highly deleterious to protein expression.

The contribution of rare polymorphisms to disease susceptibility is a growing area of interest (Tennessen et al., 2012). Therefore, I used the 1000 Genome Project to identify rare variants that are likely to cause drastic changes in the antiviral potential of Apobec3s. In addition to the 913 individuals with African, Asian, and European ancestries, I searched for rare variants in 181 individuals of South American ancestry, who were also enrolled in the 1000 Genome Project. Here, I identified a rare mutation in the start codon of Apobec3G at a frequency of 0.001 – 1.2% in Asian individuals (1000 Genome Project releases vary in their estimate) that results in a methionine to threonine (M1T) change. To confirm this SNV, we obtained B cells from one individual who is heterozygous for the M1T mutation and from two individuals without the mutation. I isolated RNA and protein from these cells and confirmed a heterozygous start codon mutation in Apobec3G cDNA from the predicted individual. Because I amplified the Apobec3G M1T allele from cellular RNA, I predicted that a truncated Apobec3G protein would be expressed in these cells. Using an antibody directed against the C-terminus of Apobec3G, I was able to measure the endogenous level of Apobec3G protein. This antibody also recognizes Apobec3A, which served as a loading control for the Western blot. Interestingly, I did not detect the truncated Apobec3G protein in the heterozygous individual, suggesting that the transcript carrying the M1T mutation does not encode a functional protein. However, after normalization to Apobec3A levels, I found that the individual who is heterozygous for the Apobec3G start codon mutation expresses approximately 60% less Apobec3G than the two individuals with intact Apobec3G start codons (Fig. 15A).

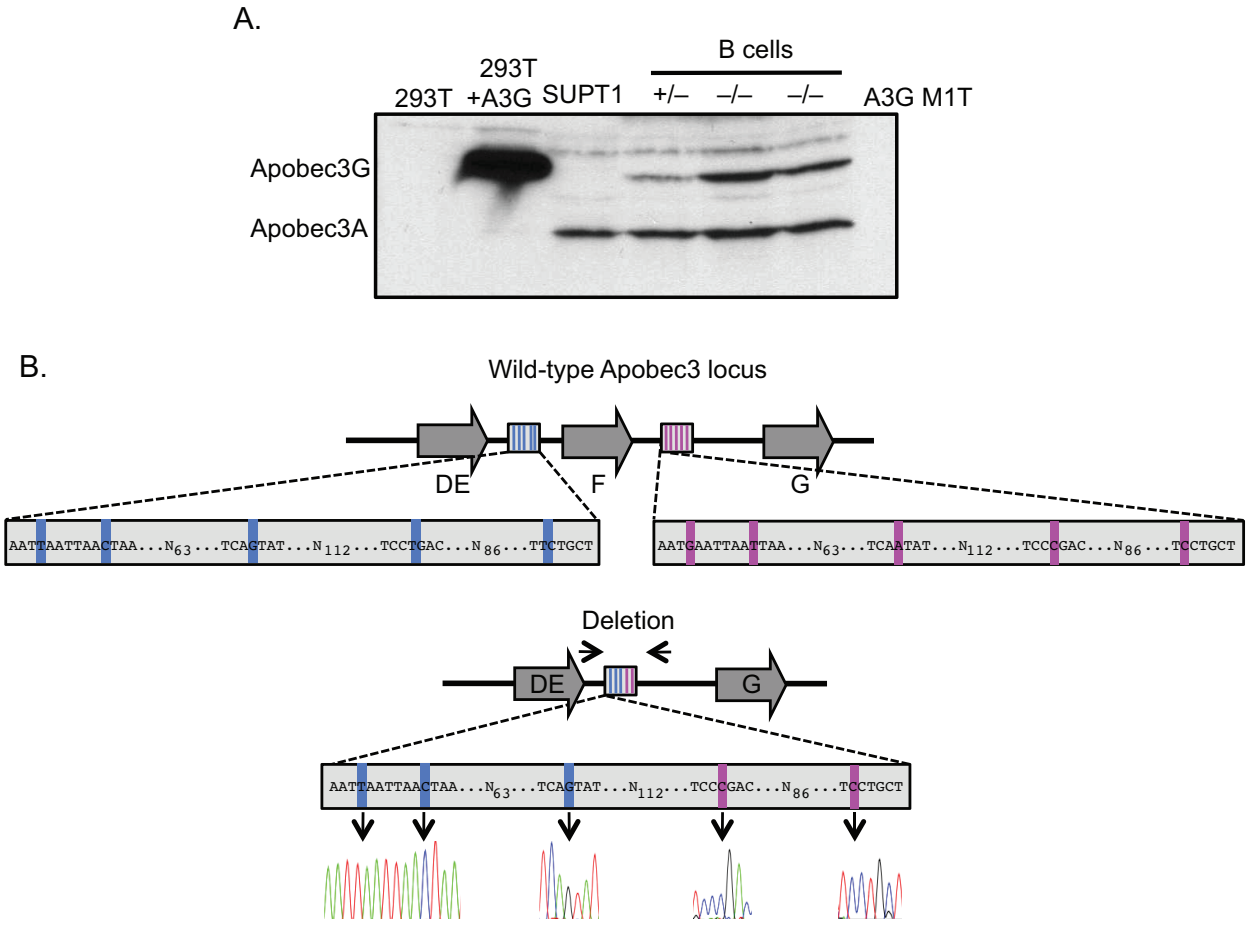


Fig. 15. Rare Apobec3 polymorphisms that are highly deleterious to protein expression.
(A) Western blot analysis of endogenous Apobec3G expression in B cell lines. Apobec3 protein was detected by Western blot using an anti-Apobec3G antibody that also recognizes Apobec3A. 293T cells transfected with Apobec3G plasmid were used as a positive control. SupT1 cells were used as a negative control. Of the 3 B cell lines, the first individual is heterozygous for the start codon mutation (A3G M1T +/–) in Apobec3G. The second two individuals do not have this mutation in Apobec3G (A3G M1T –/–). The upper band is Apobec3G, and the lower band is Apobec3A. (B) Sequencing of Apobec3F deletion in one individual. (Top) Schematic of the wild-type Apobec3 locus from the human genomic reference sequence. DNA sequences within homologous regions flanking Apobec3F are shown in gray rectangles. Sites within the homologous regions that differ are highlighted in blue (upstream of Apobec3F) and pink (downstream of Apobec3F). (Bottom) Schematic of Apobec3 locus with Apobec3F deletion. The deletion breakpoint was amplified using primers outside of Apobec3F (arrows). The PCR product was sequenced and the chromatograms are shown. The sites highlighted in blue and pink were used to confirm the novel DNA arrangement.

I also identified a large, heterozygous deletion in the Apobec3 locus of one individual of Mexican ancestry. This 27.6 kb deletion eliminates the entire Apobec3F gene (Fig. 15B, lower panel), from approximately Chr22: 39432067 – 39459664. To confirm this deletion, I isolated genomic DNA from B cells derived from this individual. Using primers flanking the deletion, I amplified a 1 kb fragment containing the breakpoint. When sequenced, the deletion was confirmed by alignments with the Apobec3 locus. I did not amplify this region from two other individuals without the deletion. As a control, I amplified a region within the Apobec3 locus from all individuals. I predicted that this deletion would decrease Apobec3F expression in this individual; however, I was unable to detect endogenous Apobec3F protein from the B cell lines of any individual using an anti-Apobec3F antibody (data not shown). Together, these data suggest that rare variants can drastically reduce Apobec3 expression.

Selection acting on the *Apobec3* locus in human populations.

By comparing the ratio of the rate of non-synonymous mutation (dN) to the rate of synonymous mutations (dS), previous studies have shown *Apobec3DE*, *Apobec3G*, and *Apobec3H* to be evolving under recurrent ancient positive selection in primates (Duggal et al., 2011; OhAinle et al., 2006; Sawyer et al., 2004). We asked if the *Apobec3* locus is also evolving under recent selective pressure within human populations and whether the deleterious mutations in Apobec3DE could have persisted as the result of neutral evolution. I calculated the interspecies divergence and intraspecies nucleotide diversity of *Apobec3A*, *C*, *DE*, *F*, *G*, and *H* for each population (African, Asian, and European ancestries). I did not evaluate *Apobec3B* because its copy number is highly variable in humans. I directly compared the synonymous and non-synonymous polymorphism within human populations to the synonymous and non-synonymous

divergence of these genes between humans and chimpanzees using the McDonald Kreitman (MK) test. Using this test, I found that *Apobec3G* has evolved under positive selection since divergence from the chimpanzee-human ancestor, though not within human populations ($p<0.05$, Table 5). To determine whether domains of *Apobec3G* are differentially evolving, I performed a sliding window analysis of polymorphism and divergence across *Apobec3G* and found that the difference between polymorphism and divergence is greatest across the 3' half of the gene, suggesting that this portion of the gene is the target of positive selection (Fig. 16A).

I interrogated the *Apobec3* genes for signatures of selection by evaluating the overall diversity and site frequency spectrum in human populations. I found that intraspecies nucleotide diversity (π) varies between *Apobec3* genes and across each gene, but diversity is similar between populations (Fig. 16A). The *Apobec3* genes with the lowest average nucleotide diversity are *Apobec3C* and *Apobec3DE* (Table 6). Using the Hudson Kreitman Aguade (HKA) test, I compared the polymorphic and fixed changes in *Apobec3* genes to the polymorphic and fixed changes in a non-coding (putatively neutral) region within the *Apobec3* locus. I found the nucleotide diversity in *Apobec3DE* in humans to be significantly lower than expected under neutrality (HKA $\chi^2 = 4.31$, $p<0.05$, Table 6). The trend is similar within each human population (African, Asian, European), suggesting that this is not a population-specific event (Table 6). This test is indicative of selective or demographic events acting on *Apobec3DE* to reduce diversity. To determine whether a selective sweep has reduced the diversity across the *Apobec3DE* genomic locus in each population, I calculated the polymorphism and divergence across *Apobec3DE* exons and introns. The nucleotide diversity in *Apobec3DE* is not consistently low across the entire genomic region, as the diversity within exons is lower than the diversity within introns (Fig. 16B, position of exons highlighted in gray). This is suggestive of purifying selection

acting on the coding regions of *Apobec3DE*, rather than a selective sweep. In addition, the Fay and Wu's H test statistic for *Apobec3DE* is not consistent with positive selection (Table 6), further suggesting that *Apobec3DE* is likely evolving under purifying selection. This could explain the deleterious R97C and R248K polymorphisms, which are present at low frequencies and may be subject to weak purifying selection.

		Silent		Replacement		Fisher's
		Polymorphic	Fixed	Polymorphic	Fixed	exact test
Apobec3A	African	5	2	2	3	n.s.
597 nt	Asian	4	2	1	4	n.s.
	European	4	2	1	4	n.s.
	Combined	5	2	3	3	n.s.
Apobec3C	African	0	1	1	4	n.s.
570 nt	Asian	0	1	2	3	n.s.
	European	0	1	0	4	n.s.
	Combined	1	1	3	3	n.s.
Apobec3DE	African	1	6	6	36	n.s.
1159 nt	Asian	0	6	3	36	n.s.
	European	0	6	4	36	n.s.
	Combined	1	6	12	35	n.s.
Apobec3F	African	5	5	5	12	n.s.
1119 nt	Asian	2	5	5	11	n.s.
	European	1	5	6	12	n.s.
	Combined	5	5	9	11	n.s.
Apobec3G	African	4	2	5	17	#
1152 nt	Asian	2	2	3	17	n.s.
	European	1	2	2	17	n.s.
	Combined	5	2	6	17	*
Apobec3H	African	3	1	8	8	n.s.
549 nt	Asian	2	1	6	8	n.s.
	European	1	1	5	8	n.s.
	Combined	3	1	8	8	n.s.

Table 5. McDonald Kreitman test of *Apobec3* genes.
*$p<0.05$, #$p<0.10$, n.s. = not significant

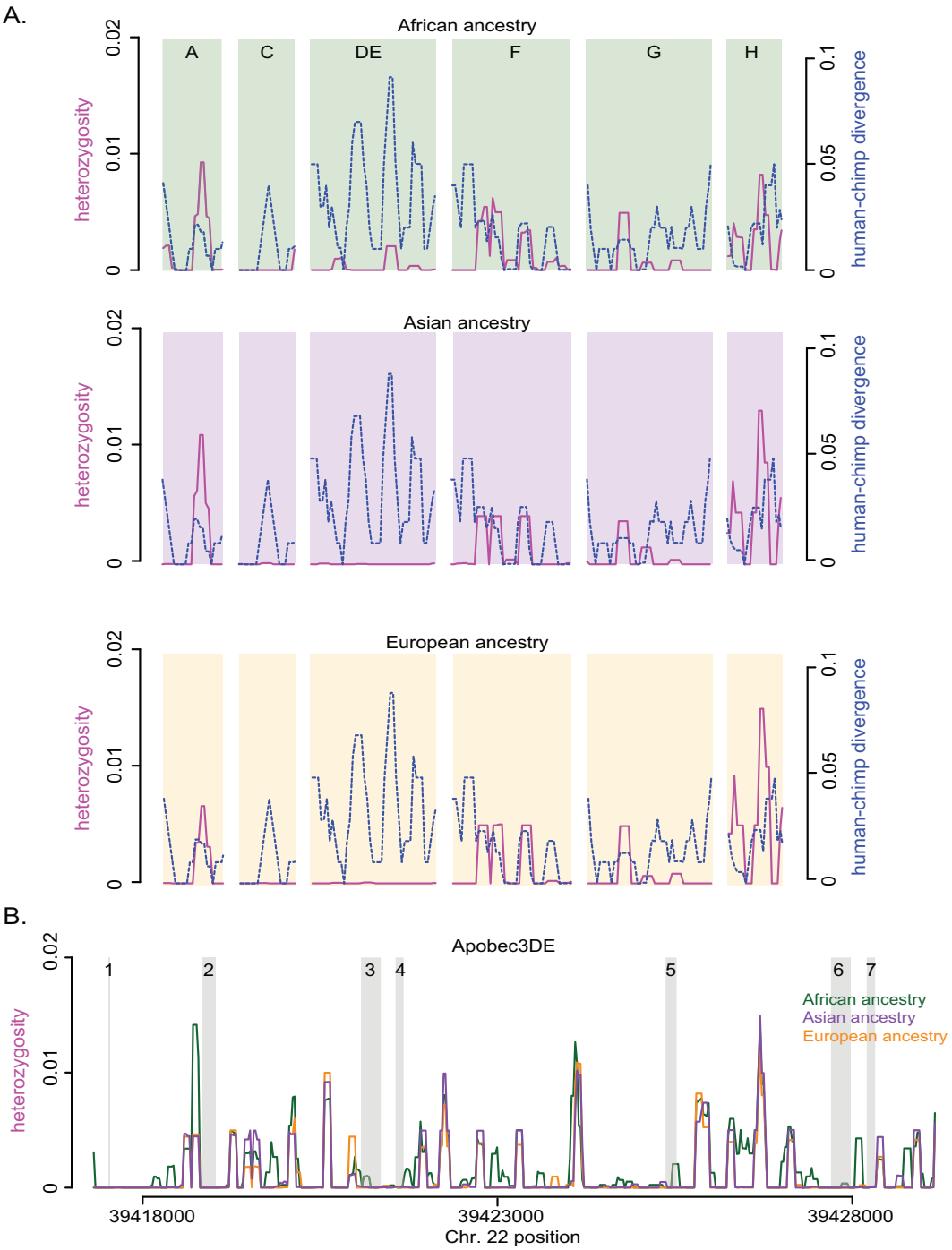


Fig. 16. Diversity across *Apobec3* genes within human populations.
Sliding window analysis of diversity, stratified by population. Green represents individuals of African ancestry, purple represents individuals of Asian ancestry, and orange represents individuals of European ancestry. (A) Coding regions of *Apobec3A*, *Apobec3C*, *Apobec3DE*, *Apobec3F*, *Apobec3G*, and *Apobec3H*. Intraspecies hterozygosity (π) is plotted in solid pink lines using the left axis, and interspecies divergence is plotted in dashed blue lines using the right axis. (B) Intronic and exonic of *Apobec3DE*. Positions of exons are highlighted in gray. Only intraspecies hterozygosity (π) is plotted for each population.

		S[a]	π[b]	Divergence[c]	Tajima's D	Fay and Wu's H	HKA χ^2
Apobec3A	African	7	$1.9x10^{-3}$	0.013	0.19	-3.6	0.20
597 nt	Asian	5	$1.9x10^{-3}$	0.013	0.92	-1.29	0.58
	European	5	$1.1x10^{-3}$	0.013	-0.06	-2.44	0.53
	Combined	8	$1.6x10^{-3}$	0.013	-0.05	-3.80	0.01
Apobec3C	African	1	$3.0x10^{-4}$	0.0089	0.15	0.15	0.87
570 nt	Asian	2	$2.0x10^{-5}$	0.0087	-1.13	-1.97	0.54
	European	1	$1.0x10^{-5}$	0.0087	-0.83	0.00	0.16
	Combined	4	$1.0x10^{-4}$	0.0088	-1.24	-1.94	0.06
Apobec3DE	African	7	$3.0x10^{-4}$	0.036	-1.27	0.31	2.40
1159 nt	Asian	3	$2.0x10^{-5}$	0.036	-1.34	0.02	2.87#
	European	4	$1.0x10^{-5}$	0.036	-1.47	-1.98	2.87#
	Combined	13	$1.0x10^{-4}$	0.036	-1.89	-1.89	4.31*
Apobec3F	African	10	$1.4x10^{-3}$	0.016	0.15	0.71	0.01
1119 nt	Asian	7	$1.2x10^{-3}$	0.016	0.55	-3.71	0.16
	European	7	$1.4x10^{-3}$	0.017	1.08	0.20	0.77
	Combined	14	$1.5x10^{-3}$	0.017	-0.13	-1.97	0.24
Apobec3G	African	9	$8.4x10^{-4}$	0.017	-0.55	0.64	0.60
1152 nt	Asian	5	$5.0x10^{-4}$	0.017	-0.34	0.45	0.19
	European	3	$5.5x10^{-4}$	0.017	0.74	0.25	0.06
	Combined	11	$7.0x10^{-4}$	0.017	-0.79	0.62	0.23
Apobec3H	African	11	$2.7x10^{-3}$	0.020	-0.18	-1.61	0.78
549 nt	Asian	8	$4.5x10^{-3}$	0.021	2.24*	-0.74	0.46
	European	6	$5.3x10^{-3}$	0.021	4.39*	0.15	0.01
	Combined	11	$5.1x10^{-3}$	0.021	2.11	0.28	0.48

Table 6. Neutrality test statistics for *Apobec3* genes.
[a] S = number of segregating sites. [b] π = nucleotide diversity. [c] Divergence = human-chimpanzee divergence. [d] HKA = Hudson Kreitman Aguade test, silent sites only. *p<0.05, #p<0.10

The highest average nucleotide diversity is seen for *Apobec3H* (Table 6). In fact, *Apobec3H* has been shown to be one of the most diverse genes in the entire genome within individuals of European lineages (Tennessen et al., 2012). The significantly positive Tajima's D test statistic for *Apobec3H* in Asian and European populations is suggestive of a population bottleneck or balancing selection acting on *Apobec3H* to maintain polymorphism in these populations (p<0.05, Table 6). This is consistent with previous observations that many genes

under balancing selection are immunity-related genes (Andrés et al., 2009). However, the HKA analysis does not support balancing selection acting on *Apobec3H*, so we cannot rule out neutral evolution.

From these population genetics analyses, I conclude that, while *Apobec3DE*, *Apobec3G*, and *Apobec3H* have previously been shown to evolve under positive selection during human-chimpanzee divergence, it is unlikely that *Apobec3* genes are evolving under positive selection within humans. Importantly, *Apobec3DE* has evolved under purifying selection in recent human history, suggesting that a cellular function of Apobec3DE has been optimized and that it is conserved due to continuous selective pressure.

A cellular role for Apobec3DE in host defense

One potential continuous selective pressure that may have driven the conservation of Apobec3DE is the Alu retrotransposon. This is because Alu elements have been active throughout human and chimpanzee evolution and are still active today (Batzer and Deininger, 2002). In addition, anti-Alu activity of Apobec3DE is conserved between human and chimpanzee Apobec3DE, despite bursts of positive selection in the chimpanzee lineage (Duggal et al., 2011). If anti-Alu activity is an essential function of Apobec3DE, we expect few polymorphisms that could decrease anti-Alu activity in Apobec3DE.

To test whether the polymorphisms in Apobec3DE affect anti-Alu activity of Apobec3DE, I compared the anti-Alu activity of wild-type Apobec3DE to the anti-Alu activity of Apobec3DE carrying the R97C or R248K mutation. Interestingly, while the R97C and R248K mutations both decreased Apobec3DE antiviral activity against HIV-1Δ*vif*, only the R248K mutation decreased Apobec3DE antiviral activity against Alu elements. At every dilution,

Apobec3DE R248K (Fig. 17, gray circles) had 4- to 5-fold lower antiviral activity than wild-type Apobec3DE, where as Apobec3DE R97C (Fig. 17, open circles) restricted Alu similarly to wild-type Apobec3DE. These data suggest that the R248K mutation in Apobec3DE causes a global decrease in antiviral activity, likely due to its low protein expression, whereas the R97C mutation decreases HIV-1, but not Alu, restriction.

Thus, anti-Alu activity is maintained in more individuals than is anti-HIV activity. If anti-Alu activity is considered as the main cellular role of Apobec3DE, there is only one low frequency deleterious polymorphism in Apobec3DE.

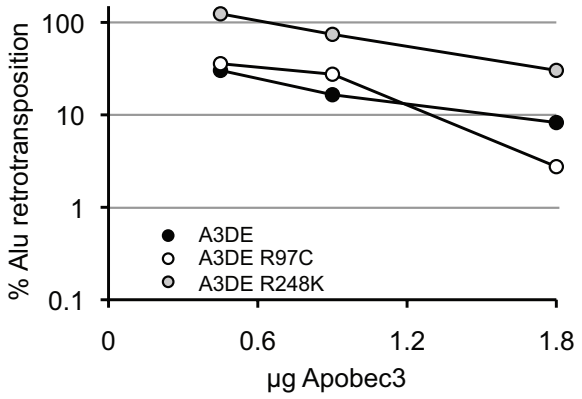


Fig. 17. Effect of SNVs on Apobec3DE restriction of Alu elements.
The antiviral activity of Apobec3DE R97C and R248K mutants against Alu elements was compared to the antiviral activity of wild-type Apobec3DE. Retrotransposition is represented as a percentage of Alu activity in the absence of Apobec3DE (A3DE), which was normalized to 100%. Wild-type Apobec3DE is represented by filled circles, Apobec3DE R97C is represented by open circles, and Apobec3DE R248K is represented by gray circles. Assays were performed with 2-fold dilutions of Apobec3DE, and Apobec3DE levels are increasing from left to right.

Discussion

In this chapter, I characterized the level of diversity in the *Apobec3* locus in human populations and the effects of variants on Apobec3 antiviral activity. I found that two low frequency variants

in Apobec3DE (R97C and R248K), which are limited to African populations, decrease the antiviral activity of Apobec3DE against HIV-1, and one variant (R248K) also decreases anti-Alu activity. I show that other rare variants identified in Apobec3G and Apobec3F drastically reduce endogenous protein expression. However, I do not find that high frequency SNVs in Apobec3A, C, F, and G genes affect antiviral activity. Further, I show that the pattern of diversity in the *Apobec3DE* is indicative of recent purifying selection in humans. Together, this data suggests that *Apobec3DE* plays an important role in host defense in humans and that low frequency and rare variants are most likely to affect the host response to currently circulating viruses.

This observation of recent purifying selection in *Apobec3DE* (Table 6 and Fig. 16) is in contrast to the ancient positive evolution that has been implicated for many Apobec3 family members. The arms race between restriction factors and retroviruses is generally thought to induce rapid evolution in immune-related genes and viruses (Duggal and Emerman, 2012). However, recent purifying selection suggests that the antiviral activity of the human repertoire of *Apobec3* genes was optimized against a selective pressure, and this antiviral activity has been maintained. While Apobec3 restriction factors have an important role in host defense in humans, the consequences of purifying selective pressure are that some Apobec3 genes, such as Apobec3DE, are not optimized to restrict the lentiviruses circulating today.

We suggest that anti-Alu activity is the critical cellular function that has driven purifying selection of Apobec3DE in humans. Because Apobec3DE has rapidly evolved in chimpanzees yet has not altered anti-Alu activity (Duggal et al., 2011), it is likely that this function is subject to purifying selection in chimpanzees as well. However, in the absence of an additional selective pressure, purifying selection is the main evolutionary force affecting Apobec3DE in humans. This is supported by previous studies showing that Apobec3DE mRNA is expressed in human

embryonic stem cells (Wissing et al., 2011), the cell type where Alu elements can have the most deleterious effect on a population. In addition, because endogenous mobile elements do not trigger an interferon response, it is interesting that Apobec3DE mRNA is constitutively expressed with only moderate upregulation by interferon. Because rare, deleterious variants are suggestive of purifying selection, the novel variants in Apobec3F and Apobec3G described here (Fig. 15) suggest that purifying selection may be acting on these Apobec3 genes as well.

However, despite the purifying selection acting on *Apobec3DE*, the R248K polymorphism does decrease anti-Alu activity (Fig. 17). Because the R248K mutation likely has decreased anti-Alu activity due to its low protein expression, its deleterious effects may be compensated *in vivo* by another polymorphism that was not included in the plasmid used for transfection. For example, the R248K polymorphism is in tight linkage disequlibrium ($r^2 > 0.8$) with two additional polymorphisms in non-coding regions near Apobec3DE that may compensate for the low *in vitro* expression level. It is also possible that the R248K polymorphism is only slightly deleterious and that the purifying selection acting on *Apobec3DE* is weak. This is supported by the low frequency of the R248K polymorphism in the human population. Either way, I conclude that *Apobec3DE* is subject to purifying selection, and the anti-Alu function of Apobec3DE is conserved in most individuals.

It has recently been shown that the majority of SNVs in human populations are rare variants, and these rare variants are predicted to be enriched for deleterious effects (Tennessen et al., 2012). By functionally testing rare variants of Apobec3 genes, my data support this prediction. Interestingly, SIFT and PolyPhen prediction programs do not predict the deleterious effect of the polymorphisms in Apobec3DE, suggesting that even more variants may be deleterious than previously thought.

Chapter V
Implications for natural HIV/SIV infection

In this work, I have characterized the evolutionary history of the *Apobec3* locus in humans and the *Apobec3DE* gene across primates. My work has shown that *Apobec3DE* has evolved rapidly in chimpanzees in response to an ancient lentivirus, and *Apobec3DE* has evolved slowly in recent human history in response to retrotransposons. I have also determined the functional effects of this evolution. Chimpanzees have gained breadth in Apobec3DE antiviral activity due to positive selection, and some human populations have lost Apobec3DE antiviral activity due to weak purifying selection. These results raise additional questions: does Apobec3DE play an *in vivo* role in modern primate innate immunity? Why are there so many deleterious mutations in Apobec3 genes in human populations? In this chapter I will discuss how the complex dynamics of the ancient and recurrent Apobec3-retrovirus interaction may have consequences for contemporary human health and viral evolution.

Contributions of Apobec3DE to innate immunity

Human Apobec3DE is very likely to have a current role in cell intrinsic immunity, as it can restrict the activity of retrotransposons that are still active in the human population (Fig. 6 and 7). Even though other human Apobec3s can restrict Alu elements, human Apobec3DE may have an expression profile that differs from other Apobec3s, thereby retaining its selective pressure and leading to its conservation as a defense factor across primates. In support of a potential role in host defense, other studies have found broad human Apobec3DE mRNA expression in many tissues, including embryonic stem cells, the cell type in which retrotransposons can exert the most damage (Dang et al., 2006; Koning et al., 2009; Refsland et al., 2010; Wissing et al., 2011).

To test this idea, I could knock-down endogenous Apobec3DE expression in an embryonic stem cell line using shRNA and perform Alu retrotransposition assays in wild-type and knock-down cells. If Alu retrotransposition activity is higher in the knock-down cells compared to wild-type cells, I would conclude that Apobec3DE plays a role in suppressing endogenous retrotransposition activity in humans.

Human Apobec3DE may also be important in the innate immune response to HIV-1 infection, despite its moderate antiviral activity compared to Apobec3G and its sensitivity to HIV-1 Vif. Here, I have shown that Apobec3DE protein is endogenously expressed in T cells and that Apobec3DE expression is moderately upregulated by IFN (Fig. 15). As I have also identified human polymorphisms that drastically decrease Apobec3DE expression and anti-HIV activity (Fig. 14), it is now possible to evaluate the association of deleterious SNVs in Apobec3DE with HIV-1 acquisition, viral load, or disease progression. To do this, I would use an existing study of HIV-1 discordant couples in Western Africa (the geographic location with the highest frequency of Apobec3DE polymorphisms) and determine the Apobec3DE genotype of HIV negative individuals. Because only 4-5% of individuals should have two deleterious Apobec3DE alleles, I would select individuals with 2 deleterious Apobec3DE alleles and match them by exposure to individuals who are homozygous for wild-type Apobec3DE. I would then test whether Apobec3DE genotype is associated with seroconversion. Within HIV positive individuals, I could also test whether Apobec3DE genotype is associated with viral load. If Apobec3DE genotype is correlated with viral transmission or replication, this would indicate that Apobec3DE antiviral activity affects HIV-1 replication *in vivo*. However, as I have only tested Apobec3DE against a single HIV-1 clone from a subtype B virus, it would also be important to show that human Apobec3DE restricts other HIV-1 subtypes, such as A and C, that are more

common in individuals from Western Africa. This would ideally be performed using primary patient viral isolates from acute infection such as Q23, the subtype A virus cloned by Julie Overbaugh (Poss and Overbaugh, 1999). Additionally, by knocking down Apobec3DE using shRNA in healthy, primary human CD4+ T cells, the impact of Apobec3DE expression on the spreading infection of HIV-1Δ*vif* could be assessed *ex vivo*. In this experiment, Apobec3G and Apobec3F expression would also have to be knocked-down, as their expression would likely prevent viral replication. These *ex vivo* experiments are promising, as a recent paper has found that knocking down Apobec3DE in a human T cell line increases HIV-1 Δ*vif* replication. (Refsland et al., 2012). To assess HIV-1 replication in the presence of Vif, multiple patient *vif* genes should be tested, as isolates vary in their strength of Apobec3 antagonism (Piantadosi et al., 2009).

As chimpanzee *Apobec3DE* rapidly evolved 2-6 million years ago, it would be interesting to evaluate *Apobec3DE* evolution in recent chimpanzee history and whether this affects restriction of SIV that infects chimpanzees (SIVcpz). The four subspecies of chimpanzees diverged approximately 500,000 – 1 million years ago (Bjork et al., 2011; Gonder et al., 2011), and only two subspecies are known to be infected with SIV (Keele et al., 2006; Leendertz et al., 2011). In the studies described here, I tested a single chimpanzee Apobec3DE allele from an unknown subspecies. I would like to sequence the Apobec3 locus from all four chimpanzee subspecies and functionally test novel haplotypes against SIVcpz. In preliminary work, we sequenced Apobec3DE from 10 *Pan troglodytes verus* (*P.t.v.*) individuals, a subpecies that is not known to be infected with SIVcpz, and found that the sequence of Apobec3DE is similar in all individuals, suggesting that Apobec3DE antiviral activity may be conserved amongst *P.t.v.* However, one chimpanzee genome contained an Apobec3DE allele with a non-synonymous

mutation in the C-terminus of the protein that decreased antiviral activity against HIV-1Δ*vif* by ten fold (data not shown). Ultimately, by functionally characterizing the chimpanzee Apobec3 repertoire, it might be possible to correlate Apobec3DE antiviral activity with SIVcpz prevalence. This might shed light on whether Apobec3DE is relevant to viral transmission amongst chimpanzees.

In addition, determining the antiviral activity of Apobec3DE from other primate species might show whether Apobec3DE has played an important role in innate immunity throughout primate history. In work not presented here, I attempted to functionally test gorilla Apobec3DE, but its poor expression in human cells prevented these experiments. However, other groups have shown that rhesus macaque Apobec3DE restricts HIV-1Δ*vif*, suggesting that Apobec3DE antiviral activity is conserved at least across hominoids and Old world monkeys (Hultquist et al., 2011). To test this idea, I would test the antiviral activity of multiple primate Apobec3DE orthologues against the SIV that naturally infects each species. For example, I would test the restriction of SIVmac by rhesus macaque Apobec3DE. If rhesus macaque Apobec3DE can restrict SIVmacΔ*vif* and SIVmac Vif antagonizes Apobec3DE, then I would conclude that Apobec3DE plays a role in innate immunity in rhesus macaques.

Viral adaptation to Apobec3 genes

In my work, I have investigated the evolution of restriction factors in order to understand how viruses shape the innate immune system. A more complete description of the host-virus interaction would require evaluating the viral adaptation to the host. For example, if Apobec3DE does present a block to natural HIV-1 infection, the virus must evolve to antagonize Apobec3DE. To show this, a longitudinal analysis of HIV-1 evolution could be performed. I would genotype

individuals from an HIV-positive cohort, select individuals with two deleterious Apobec3DE alleles, and match them by duration of infection to individuals who are homozygous for wild-type Apobec3DE. I predict that HIV-1 from individuals with deleterious Apobec3DE alleles would have lower levels of hypermutation and might lose the ability to antagonize Apobec3DE over time.

In addition, evaluating the sequence of Vif before and after cross-species transmissions could identify viral adaptations to Apobec3 genes. Here, I showed that human and chimpanzee Apobec3DE are sensitive to Vif from HIV-1 and SIV that infects *P.t. schweinfurthii* (*P.t.s.*) (Fig. 8), but I did not test their sensitivity to Vif from SIV that infects *P.t. troglodytes* (*P.t.t.*), which is the SIV lineage that led to the HIV-1 pandemic in humans. I would clone multiple *vif* genes from SIVcpz*Ptt* and HIV-1 and test their ability to antagonize all common human and chimpanzee Apobec3 haplotypes to determine whether SIVcpz Vif evolved during cross-species transmission to humans. In addition, if there are differences in Vif antagonism of Apobec3s between SIVcpz*Ptt* and SIVcpz*Pts*, it is possible that Apobec3s may have contributed to the cross-species transmission potential of these two lineages of SIVcpz.

Why are there so many deleterious alleles in the Apobec3 locus?

Because several *Apobec3* genes have been under strong positive selection for millions of years in primates, I anticipated that many *Apobec3* genes would also be under strong positive selection in recent human history. However, *Apobec3* genes do not have signatures of positive selection in human populations. In addition, there are many common (MAF >1%) polymorphisms in *Apobec3* genes that decrease protein expression and/or antiviral activity: two polymorphisms in Apobec3DE (R97C and R248K), one polymorphism in Apobec3F (Y307C), two variants of

Apobec3H (R105G and the deletion of residue 15), and the deletion of Apobec3B. Whether this level of deleterious alleles is typical or unusual for a human gene is unclear. However, one possible reason for the maintenance of these deleterious alleles in our innate immune system is that, just as the *Apobec3* gene family has expanded in primates, the *Apobec3* locus is contracting in humans. If Apobec3s impose a cost on the host, we would expect that the loss of *Apobec3* genes would be selected for in the absence of a viral infection. In this case, I would expect signatures of positive selection surrounding the deleterious allele, but, in my analyses, I do not see that deleterious alleles are the target of selection. Another possibility is that *Apobec3* genes evolve under different types of selective pressure in short time-scales compared to long time-scales. For example, many genes involved in host defense have been shown to be evolving under balancing selection within human populations. This may be advantageous to the host, as host heterozygosity may broaden the host defense system to recognize multiple pathogens, or it may prevent efficient viral adaptation to the host. In my analyses, one gene containing deleterious alleles does have signatures of balancing selection (*Apobec3H*). If multiple alleles of *Apobec3H* were maintained in human populations, I would expect each haplotype to contain signatures of selection, such as extended haplotype homozygosity. A third possibility is that the population size of humans has expanded so dramatically in recent history that deleterious alleles have not yet been selected out of the population. In this case, we would expect these deleterious polymorphisms to be recently acquired in humans. To test this, I would perform a coalescent analysis to estimate the time to most recent common ancestor (TMRCA) of the deleterious alleles.

In any case, my work suggests that the functional effect of genetic variation in human populations needs to be characterized experimentally. Many studies are interested in identifying

the genetic determinants of human health conditions, but few studies have found that common genetic variants can explain complex diseases. There is now a greater emphasis on associating rare genetic variants with disease. However, whether this will be statistically feasible is still unknown, and it is possible that only *in vitro* functional experiments will allow the identification of mutations that affect disease outcome. It is my hope that the work presented here can be an example of this type of analysis and can contribute to future efforts in determining the genetic component of human susceptibility to viral infection.

By studying the evolution of host-virus interactions on a molecular level, I have demonstrated that host restriction factors have evolved rapidly in response to previous viral infections. Further studies will define how restriction factors play a role in extant viral infections. Ultimately, the innate immune response to contemporary circulating viruses and future viral zoonoses are the result of an ancient and ongoing arms race between the host and virus.

References

Aguiar, R.S., Lovsin, N., Tanuri, A., and Peterlin, B.M. (2008). Vpr.A3A chimera inhibits HIV replication. J Biol Chem *283*, 2518-2525.

Albin, J.S., and Harris, R.S. (2010). Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics. Expert Rev Mol Med *12*, e4.

Alisch, R.S., Garcia-Perez, J.L., Muotri, A.R., Gage, F.H., and Moran, J.V. (2006). Unconventional translation of mammalian LINE-1 retrotransposons. Genes Dev *20*, 210-224.

An, P., Bleiber, G., Duggal, P., Nelson, G., May, M., Mangeat, B., Alobwede, I., Trono, D., Vlahov, D., Donfield, S.*, et al.* (2004). APOBEC3G genetic variants and their influence on the progression to AIDS. J Virol *78*, 11070-11076.

Andrés, A.M., Hubisz, M.J., Indap, A., Torgerson, D.G., Degenhardt, J.D., Boyko, A.R., Gutenkunst, R.N., White, T.J., Green, E.D., Bustamante, C.D.*, et al.* (2009). Targets of balancing selection in the human genome. Mol Biol Evol *26*, 2755-2764.

Barrett, B.S., Smith, D.S., Li, S.X., Guo, K., Hasenkrug, K.J., and Santiago, M.L. (2012). A single nucleotide polymorphism in tetherin promotes retrovirus restriction in vivo. PLoS Pathog *8*, e1002596.

Bartz, S.R., and Vodicka, M.A. (1997). Production of high-titer human immunodeficiency virus type 1 pseudotyped with vesicular stomatitis virus glycoprotein. Methods *12*, 337-342.

Batzer, M.A., and Deininger, P.L. (2002). Alu repeats and human genomic diversity. Nat Rev Genet *3*, 370-379.

Bennett, E.A., Keller, H., Mills, R.E., Schmidt, S., Moran, J.V., Weichenrieder, O., and Devine, S.E. (2008). Active Alu retrotransposons in the human genome. Genome Res *18*, 1875-1883.

Bishop, K.N., Holmes, R.K., Sheehy, A.M., Davidson, N.O., Cho, S.J., and Malim, M.H. (2004). Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Curr Biol *14*, 1392-1396.

Bishop, K.N., Verma, M., Kim, E.Y., Wolinsky, S.M., and Malim, M.H. (2008). APOBEC3G inhibits elongation of HIV-1 reverse transcripts. PLoS Pathog *4*, e1000231.

Biswas, S., and Akey, J.M. (2006). Genomic insights into positive selection. Trends Genet *22*, 437-446.

Bjork, A., Liu, W., Wertheim, J.O., Hahn, B.H., and Worobey, M. (2011). Evolutionary history of chimpanzees inferred from complete mitochondrial genomes. Mol Biol Evol *28*, 615-623.

Bogerd, H.P., Doehle, B.P., Wiegand, H.L., and Cullen, B.R. (2004). A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor. Proc Natl Acad Sci U S A *101*, 3770-3774.

Bogerd, H.P., Tallmadge, R.L., Oaks, J.L., Carpenter, S., and Cullen, B.R. (2008). Equine infectious anemia virus resists the antiretroviral activity of equine APOBEC3 proteins through a packaging-independent mechanism. J Virol *82*, 11889-11901.

Bogerd, H.P., Wiegand, H.L., Doehle, B.P., Lueders, K.K., and Cullen, B.R. (2006a). APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. Nucleic Acids Res *34*, 89-95.

Bogerd, H.P., Wiegand, H.L., Hulme, A.E., Garcia-Perez, J.L., O'Shea, K.S., Moran, J.V., and Cullen, B.R. (2006b). Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci U S A *103*, 8780-8785.

Brennan, G., Kozyrev, Y., and Hu, S.L. (2008). TRIMCyp expression in Old World primates Macaca nemestrina and Macaca fascicularis. Proc Natl Acad Sci U S A *105*, 3569-3574.

Cen, S., Peng, Z.G., Li, X.Y., Li, Z.R., Ma, J., Wang, Y.M., Fan, B., You, X.F., Wang, Y.P., Liu, F.*, et al.* (2010). Small molecular compounds inhibit HIV-1 replication through specifically stabilizing APOBEC3G. J Biol Chem *285*, 16546-16552.

Chen, H., Lilley, C.E., Yu, Q., Lee, D.V., Chou, J., Narvaiza, I., Landau, N.R., and Weitzman, M.D. (2006). APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. Curr Biol *16*, 480-485.

Chiu, Y.L., Soros, V.B., Kreisberg, J.F., Stopak, K., Yonemoto, W., and Greene, W.C. (2005). Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells. Nature *435*, 108-114.

Compton, A.A., Hirsch, V.M., and Emerman, M. (2012). The host restriction factor APOBEC3G and retroviral Vif protein coevolve due to ongoing genetic conflict. Cell Host Microbe *11*, 91-98.

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

Dang, Y., Abudu, A., Son, S., Harjes, E., Spearman, P., Matsuo, H., and Zheng, Y.H. (2011). Identification of a single amino acid required for APOBEC3 antiretroviral cytidine deaminase activity. J Virol *85*, 5691-5695.

Dang, Y., Wang, X., Esselman, W.J., and Zheng, Y.H. (2006). Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. J Virol *80*, 10522-10533.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M.*, et al.* (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res *36*, W465-469.

Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. Nat Genet *35*, 41-48.

Doehle, B.P., Schafer, A., and Cullen, B.R. (2005). Human APOBEC3B is a potent inhibitor of HIV-1 infectivity and is resistant to HIV-1 Vif. Virology *339*, 281-288.

Duggal, N.K., and Emerman, M. (2012). Evolutionary conflicts between viruses and restriction factors shape immunity. Nat Rev Immunol.

Duggal, N.K., Malik, H.S., and Emerman, M. (2011). The breadth of antiviral activity of Apobec3DE in chimpanzees has been driven by positive selection. J Virol *85*, 11361-11371.

Elde, N.C., and Malik, H.S. (2009). The evolutionary conundrum of pathogen mimicry. Nat Rev Microbiol *7*, 787-797.

Emerman, M., and Malik, H.S. (2010). Paleovirology--modern consequences of ancient viruses. PLoS Biol *8*, e1000301.

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

Everitt, A.R., Clare, S., Pertel, T., John, S.P., Wash, R.S., Smith, S.E., Chin, C.R., Feeley, E.M., Sims, J.S., Adams, D.J.*, et al.* (2012). IFITM3 restricts the morbidity and mortality associated with influenza. Nature *484*, 519-523.

Ganser-Pornillos, B.K., Chandrasekaran, V., Pornillos, O., Sodroski, J.G., Sundquist, W.I., and Yeager, M. (2011). Hexagonal assembly of a restricting TRIM5alpha protein. Proc Natl Acad Sci U S A *108*, 534-539.

Gifford, R.J., Katzourakis, A., Tristem, M., Pybus, O.G., Winters, M., and Shafer, R.W. (2008). A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. Proc Natl Acad Sci U S A *105*, 20362-20367.

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

Goila-Gaur, R., Khan, M.A., Miyagi, E., Kao, S., and Strebel, K. (2007). Targeting APOBEC3A to the viral nucleoprotein complex confers antiviral activity. Retrovirology *4*, 61.

Gonder, M.K., Locatelli, S., Ghobrial, L., Mitchell, M.W., Kujawski, J.T., Lankester, F.J., Stewart, C.B., and Tishkoff, S.A. (2011). Evidence from Cameroon reveals differences in the genetic structure and histories of chimpanzee populations. Proc Natl Acad Sci U S A *108*, 4766-4771.

Gooch, B.D., and Cullen, B.R. (2008). Functional domain organization of human APOBEC3G. Virology *379*, 118-124.

Haller, O., Acklin, M., and Staeheli, P. (1987). Influenza virus resistance of wild mice: wild-type and mutant Mx alleles occur at comparable frequencies. J Interferon Res *7*, 647-656.

Han, G.-Z., and Worobey, M. (2012). An Endogenous Foamy-like Viral Element in the Coelacanth Genome. PLoS Pathog *8*, e1002790.

Hanada, K., Suzuki, Y., and Gojobori, T. (2004). A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. Mol Biol Evol *21*, 1074-1080.

Harari, A., Ooms, M., Mulder, L.C., and Simon, V. (2009). Polymorphisms and splice variants influence the antiretroviral activity of human APOBEC3H. J Virol *83*, 295-303.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., and Malim, M.H. (2003). DNA deamination mediates innate immunity to retroviral infection. Cell *113*, 803-809.

Hatziioannou, T., Princiotta, M., Piatak, M., Yuan, F., Zhang, F., Lifson, J.D., and Bieniasz, P.D. (2006). Generation of simian-tropic HIV-1 by restriction factor evasion. Science *314*, 95.

Hedges, D.J., Callinan, P.A., Cordaux, R., Xing, J., Barnes, E., and Batzer, M.A. (2004). Differential alu mobilization and polymorphism among the human and chimpanzee lineages. Genome Res *14*, 1068-1075.

Holmes, E.C. (2003). Error thresholds and the constraints to RNA virus evolution. Trends Microbiol *11*, 543-546.

Hrecka, K., Hao, C., Gierszewska, M., Swanson, S.K., Kesik-Brodacka, M., Srivastava, S., Florens, L., Washburn, M.P., and Skowronski, J. (2011). Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. Nature *474*, 658-661.

Hughes, A.L., and Yeager, M. (1998). Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet *32*, 415-435.

Hulme, A.E., Bogerd, H.P., Cullen, B.R., and Moran, J.V. (2007). Selective inhibition of Alu retrotransposition by APOBEC3G. Gene *390*, 199-205.

Hultquist, J.F., Lengyel, J.A., Refsland, E.W., LaRue, R.S., Lackey, L., Brown, W.L., and Harris, R.S. (2011). Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. J Virol *85*, 11220-11234.

Huthoff, H., and Malim, M.H. (2007). Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation. J Virol *81*, 3807-3815.

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

Jenkins, G.M., Rambaut, A., Pybus, O.G., and Holmes, E.C. (2002). Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J Mol Evol *54*, 156-165.

Johnson, W.E., and Sawyer, S.L. (2009). Molecular evolution of the antiretroviral TRIM5 gene. Immunogenetics *61*, 163-176.

Kaiser, S.M., and Emerman, M. (2006). Uracil DNA glycosylase is dispensable for human immunodeficiency virus type 1 replication and does not contribute to the antiviral effects of the cytidine deaminase Apobec3G. J Virol *80*, 875-882.

Kamada, K., Igarashi, T., Martin, M.A., Khamsri, B., Hatcho, K., Yamashita, T., Fujita, M., Uchiyama, T., and Adachi, A. (2006). Generation of HIV-1 derivatives that productively infect macaque monkey lymphoid cells. Proc Natl Acad Sci U S A *103*, 16959-16964.

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

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

Keele, B.F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M.L., Bibollet-Ruche, F., Chen, Y., Wain, L.V., Liegeois, F*., et al.* (2006). Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science *313*, 523-526.

Kidd, J.M., Newman, T.L., Tuzun, E., Kaul, R., and Eichler, E.E. (2007). Population stratification of a common APOBEC gene deletion polymorphism. PLoS Genet *3*, e63.

Kirchhoff, F. (2010). Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses. Cell Host Microbe *8*, 55-67.

Kirmaier, A., Wu, F., Newman, R.M., Hall, L.R., Morgan, J.S., O'Connor, S., Marx, P.A., Meythaler, M., Goldstein, S., Buckler-White, A*., et al.* (2010). TRIM5 suppresses cross-species transmission of a primate immunodeficiency virus and selects for emergence of resistant variants in the new species. PLoS Biol *8*.

Koning, F.A., Newman, E.N., Kim, E.Y., Kunstman, K.J., Wolinsky, S.M., and Malim, M.H. (2009). Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. J Virol *83*, 9474-9485.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., and Frost, S.D. (2006). Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol *23*, 1891-1901.

Kueck, T., and Neil, S.J. (2012). A cytoplasmic tail determinant in HIV-1 Vpu mediates targeting of tetherin for endosomal degradation and counteracts interferon-induced restriction. PLoS Pathog *8*, e1002609.

Laguette, N., Rahm, N., Sobhian, B., Chable-Bessia, C., Münch, J., Snoeck, J., Sauter, D., Switzer, W.M., Heneine, W., Kirchhoff, F.*, et al.* (2012). Evolutionary and functional analyses of the interaction between the myeloid restriction factor SAMHD1 and the lentiviral Vpx protein. Cell Host Microbe *11*, 205-217.

Laguette, N., Sobhian, B., Casartelli, N., Ringeard, M., Chable-Bessia, C., Ségéral, E., Yatim, A., Emiliani, S., Schwartz, O., and Benkirane, M. (2011). SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. Nature *474*, 654-657.

Lahouassa, H., Daddacha, W., Hofmann, H., Ayinde, D., Logue, E.C., Dragin, L., Bloch, N., Maudet, C., Bertrand, M., Gramberg, T.*, et al.* (2012). SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. Nat Immunol *13*, 223-228.

LaRue, R.S., Jónsson, S.R., Silverstein, K.A., Lajoie, M., Bertrand, D., El-Mabrouk, N., Hötzel, I., Andrésdóttir, V., Smith, T.P., and Harris, R.S. (2008). The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. BMC Mol Biol *9*, 104.

Le Tortorec, A., Willey, S., and Neil, S.J. (2011). Antiviral inhibition of enveloped virus release by tetherin/BST-2: action and counteraction. Viruses *3*, 520-540.

Leendertz, S.A., Locatelli, S., Boesch, C., Kücherer, C., Formenty, P., Liegeois, F., Ayouba, A., Peeters, M., and Leendertz, F.H. (2011). No evidence for transmission of SIVwrc from western red colobus monkeys (Piliocolobus badius badius) to wild West African chimpanzees (Pan troglodytes verus) despite high exposure through hunting. BMC Microbiol *11*, 24.

Li, M.M., Wu, L.I., and Emerman, M. (2010). The range of human APOBEC3H sensitivity to lentiviral Vif proteins. J Virol *84*, 88-95.

Liao, C.H., Kuang, Y.Q., Liu, H.L., Zheng, Y.T., and Su, B. (2007). A novel fusion gene, TRIM5-Cyclophilin A in the pig-tailed macaque determines its susceptibility to HIV-1 infection. AIDS *21 Suppl 8*, S19-26.

Liberatore, R.A., and Bieniasz, P.D. (2011). Tetherin is a key effector of the antiretroviral activity of type I interferon in vitro and in vivo. Proc Natl Acad Sci U S A *108*, 18097-18101.

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics *25*, 1451-1452.

Liddament, M.T., Brown, W.L., Schumacher, A.J., and Harris, R.S. (2004). APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. Curr Biol *14*, 1385-1391.

Lilly, F. (1967). Susceptibility to two strains of Friend leukemia virus in mice. Science *155*, 461-462.

Lim, E.S., Fregoso, O.I., McCoy, C.O., Matsen, F.A., Malik, H.S., and Emerman, M. (2012). The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. Cell Host Microbe *11*, 194-204.

MacGinnitie, A.J., Anant, S., and Davidson, N.O. (1995). Mutagenesis of apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, reveals distinct domains that mediate cytosine nucleoside deaminase, RNA binding, and RNA editing activity. J Biol Chem *270*, 14768-14775.

Malim, M.H., and Emerman, M. (2008). HIV-1 accessory proteins--ensuring viral survival in a hostile environment. Cell Host Microbe *3*, 388-398.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature *424*, 99-103.

Mangeat, B., Turelli, P., Liao, S., and Trono, D. (2004). A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action. J Biol Chem *279*, 14481-14483.

Mariani, R., Chen, D., Schrofelbauer, B., Navarro, F., Konig, R., Bollman, B., Munk, C., Nymark-McMahon, H., and Landau, N.R. (2003). Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. Cell *114*, 21-31.

Mehle, A., Goncalves, J., Santa-Marta, M., McPike, M., and Gabuzda, D. (2004). Phosphorylation of a novel SOCS-box regulates assembly of the HIV-1 Vif-Cul5 complex that promotes APOBEC3G degradation. Genes Dev *18*, 2861-2866.

Mehle, A., Thomas, E.R., Rajendran, K.S., and Gabuzda, D. (2006). A zinc-binding region in Vif binds Cul5 and determines cullin selection. J Biol Chem *281*, 17259-17265.

Meiering, C.D., and Linial, M.L. (2001). Historical perspective of foamy virus epidemiology and infection. Clin Microbiol Rev *14*, 165-176.

Mills, R.E., Bennett, E.A., Iskow, R.C., Luttig, C.T., Tsui, C., Pittard, W.S., and Devine, S.E. (2006). Recently mobilized transposons in the human and chimpanzee genomes. Am J Hum Genet *78*, 671-679.

Moran, J.V., DeBerardinis, R.J., and Kazazian, H.H. (1999). Exon shuffling by L1 retrotransposition. Science *283*, 1530-1534.

Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Löwer, J., Cichutek, K., Flory, E., Schumann, G.G., and Münk, C. (2006). APOBEC3 proteins inhibit human LINE-1 retrotransposition. J Biol Chem *281*, 22161-22172.

Mulder, L.C., Ooms, M., Majdak, S., Smedresman, J., Linscheid, C., Harari, A., Kunz, A., and Simon, V. (2010). Moderate influence of human APOBEC3F on HIV-1 replication in primary lymphocytes. J Virol *84*, 9613-9617.

Munk, C., Beck, T., Zielonka, J., Hotz-Wagenblatt, A., Chareza, S., Battenberg, M., Thielebein, J., Cichutek, K., Bravo, I.G., O'Brien, S.J., *et al.* (2008). Functions, structure, and read-through alternative splicing of feline APOBEC3 genes. Genome Biol *9*, R48.

Nathans, R., Cao, H., Sharova, N., Ali, A., Sharkey, M., Stranska, R., Stevenson, M., and Rana, T.M. (2008). Small-molecule inhibition of HIV-1 Vif. Nat Biotechnol *26*, 1187-1192.

Newman, R.M., Hall, L., Kirmaier, A., Pozzi, L.A., Pery, E., Farzan, M., O'Neil, S.P., and Johnson, W. (2008). Evolution of a TRIM5-CypA splice isoform in old world monkeys. PLoS Pathog *4*, e1000003.

Nielsen, R. (2005). Molecular signatures of natural selection. Annu Rev Genet *39*, 197-218.

Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., and Clark, A.G. (2007). Recent and ongoing selection in the human genome. Nat Rev Genet *8*, 857-868.

Nielsen, R., and Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics *148*, 929-936.

Niewiadomska, A.M., Tian, C., Tan, L., Wang, T., Sarkis, P.T., and Yu, X.F. (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*, 9577-9583.

OhAinle, M., Kerns, J.A., Li, M.M., Malik, H.S., and Emerman, M. (2008). Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. Cell Host Microbe *4*, 249-259.

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

Okeoma, C.M., Lovsin, N., Peterlin, B.M., and Ross, S.R. (2007). APOBEC3 inhibits mouse mammary tumour virus replication in vivo. Nature *445*, 927-930.

Okeoma, C.M., Petersen, J., and Ross, S.R. (2009). Expression of murine APOBEC3 alleles in different mouse strains and their effect on mouse mammary tumor virus infection. J Virol *83*, 3029-3038.

Pertel, T., Hausmann, S., Morger, D., Züger, S., Guerra, J., Lascano, J., Reinhard, C., Santoni, F.A., Uchil, P.D., Chatel, L., *et al.* (2011). TRIM5 is an innate immune sensor for the retrovirus capsid lattice. Nature *472*, 361-365.

Piantadosi, A., Humes, D., Chohan, B., McClelland, R.S., and Overbaugh, J. (2009). Analysis of the percentage of human immunodeficiency virus type 1 sequences that are hypermutated and markers of disease progression in a longitudinal cohort, including one individual with a partially defective Vif. J Virol *83*, 7805-7814.

Pincus, T., Rowe, W.P., and Lilly, F. (1971). A major genetic locus affecting resistance to infection with murine leukemia viruses. II. Apparent identity to a major locus described for resistance to friend murine leukemia virus. J Exp Med *133*, 1234-1241.

Poss, M., and Overbaugh, J. (1999). Variants from the diverse virus population identified at seroconversion of a clade A human immunodeficiency virus type 1-infected woman have distinct biological properties. J Virol *73*, 5255-5264.

Pritchard, J.K., Pickrell, J.K., and Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol *20*, R208-215.

Reddy, K., Winkler, C.A., Werner, L., Mlisana, K., Abdool Karim, S.S., Ndung'u, T., and Team, C.A.I.S. (2010). APOBEC3G expression is dysregulated in primary HIV-1 infection and polymorphic variants influence CD4+ T-cell counts and plasma viral load. AIDS *24*, 195-204.

Refsland, E.W., Hultquist, J.F., and Harris, R.S. (2012). Endogenous Origins of HIV-1 G-to-A Hypermutation and Restriction in the Nonpermissive T Cell Line CEM2n. PLoS Pathog *8*, e1002800.

Refsland, E.W., Stenglein, M.D., Shindo, K., Albin, J.S., Brown, W.L., and Harris, R.S. (2010). Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. Nucleic Acids Res *38*, 4274-4284.

Ribet, D., Dewannieux, M., and Heidmann, T. (2004). An active murine transposon family pair: retrotransposition of "master" MusD copies and ETn trans-mobilization. Genome Res *14*, 2261-2267.

Rice, G.I., Bond, J., Asipu, A., Brunette, R.L., Manfield, I.W., Carr, I.M., Fuller, J.C., Jackson, R.M., Lamb, T., Briggs, T.A.*, et al.* (2009). Mutations involved in Aicardi-Goutières syndrome implicate SAMHD1 as regulator of the innate immune response. Nat Genet *41*, 829-832.

Russell, R.A., and Pathak, V.K. (2007). Identification of two distinct human immunodeficiency virus type 1 Vif determinants critical for interactions with human APOBEC3G and APOBEC3F. J Virol *81*, 8201-8210.

Sambrook, J., and Russell, D.W. (2006). Rapid Amplification of 3' cDNA Ends (3'-RACE). CSH Protoc *2006*.

Sauter, D., Hué, S., Petit, S.J., Plantier, J.C., Towers, G.J., Kirchhoff, F., and Gupta, R.K. (2011). HIV-1 Group P is unable to antagonize human tetherin by Vpu, Env or Nef. Retrovirology *8*, 103.

Sauter, D., Schindler, M., Specht, A., Landford, W.N., Münch, J., Kim, K.A., Votteler, J., Schubert, U., Bibollet-Ruche, F., Keele, B.F.*, et al.* (2009). Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains. Cell Host Microbe *6*, 409-421.

Sawyer, S.L., Emerman, M., and Malik, H.S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS Biol *2*, E275.

Sawyer, S.L., Emerman, M., and Malik, H.S. (2007). Discordant Evolution of the Adjacent Antiretroviral Genes TRIM22 and TRIM5 in Mammals. PLoS Pathog *3*, e197.

Schafer, A., Bogerd, H.P., and Cullen, B.R. (2004). Specific packaging of APOBEC3G into HIV-1 virions is mediated by the nucleocapsid domain of the gag polyprotein precursor. Virology *328*, 163-168.

Schrofelbauer, B., Chen, D., and Landau, N.R. (2004). A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). Proc Natl Acad Sci U S A *101*, 3927-3932.

Shackelton, L.A., and Holmes, E.C. (2006). Phylogenetic evidence for the rapid evolution of human B19 erythrovirus. J Virol *80*, 3666-3669.

Shackelton, L.A., Parrish, C.R., Truyen, U., and Holmes, E.C. (2005). High rate of viral evolution associated with the emergence of carnivore parvovirus. Proc Natl Acad Sci U S A *102*, 379-384.

Sharp, P.M., Bailes, E., Stevenson, M., Emerman, M., and Hahn, B.H. (1996). Gene acquisition in HIV and SIV. Nature *383*, 586-587.

Sharp, P.M., and Hahn, B.H. (2011). Origins of HIV and the AIDS Pandemic. Cold Spring Harb Perspect Med *1*, a006841.

Siegrist, F., Ebeling, M., and Certa, U. (2011). The small interferon-induced transmembrane genes and proteins. J Interferon Cytokine Res *31*, 183-197.

Simon, V., Zennou, V., Murray, D., Huang, Y., Ho, D.D., and Bieniasz, P.D. (2005). Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. PLoS Pathog *1*, e6.

Smith, J.L., and Pathak, V.K. (2010). Identification of specific determinants of human APOBEC3F, APOBEC3C, and APOBEC3DE and African green monkey APOBEC3F that interact with HIV-1 Vif. J Virol *84*, 12599-12608.

Sodora, D.L., Allan, J.S., Apetrei, C., Brenchley, J.M., Douek, D.C., Else, J.G., Estes, J.D., Hahn, B.H., Hirsch, V.M., Kaur, A.*, et al.* (2009). Toward an AIDS vaccine: lessons from natural simian immunodeficiency virus infections of African nonhuman primate hosts. Nat Med *15*, 861-865.

Staeheli, P., and Sutcliffe, J.G. (1988). Identification of a second interferon-regulated murine Mx gene. Mol Cell Biol *8*, 4524-4528.

Stenglein, M.D., and Harris, R.S. (2006). APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. J Biol Chem *281*, 16837-16841.

Stetson, D.B. (2009). Connections between antiviral defense and autoimmunity. Curr Opin Immunol *21*, 244-250.

Stetson, D.B., Ko, J.S., Heidmann, T., and Medzhitov, R. (2008). Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell *134*, 587-598.

Swiecki, M., Wang, Y., Gilfillan, S., Lenschow, D.J., and Colonna, M. (2012). Cutting edge: paradoxical roles of BST2/tetherin in promoting type I IFN response and viral infection. J Immunol *188*, 2488-2492.

Switzer, W.M., Salemi, M., Shanmugam, V., Gao, F., Cong, M.E., Kuiken, C., Bhullar, V., Beer, B.E., Vallet, D., Gautier-Hion, A.*, et al.* (2005). Ancient co-speciation of simian foamy viruses and primates. Nature *434*, 376-380.

Takaoka, A., and Yanai, H. (2006). Interferon signalling network in innate defence. Cell Microbiol *8*, 907-922.

Takeda, E., Tsuji-Kawahara, S., Sakamoto, M., Langlois, M.A., Neuberger, M.S., Rada, C., and Miyazawa, M. (2008). Mouse APOBEC3 restricts friend leukemia virus infection and pathogenesis in vivo. J Virol *82*, 10998-11008.

Tan, K.S., Olfat, F., Phoon, M.C., Hsu, J.P., Howe, J.L., Seet, J.E., Chin, K.C., and Chow, V.T. (2012). In vivo and in vitro studies on the antiviral activities of viperin against influenza H1N1 virus infection. J Gen Virol *93*, 1269-1277.

Tan, L., Sarkis, P.T., Wang, T., Tian, C., and Yu, X.F. (2009). Sole copy of Z2-type human cytidine deaminase APOBEC3H has inhibitory activity against retrotransposons and HIV-1. FASEB J *23*, 279-287.

Tareen, S.U., and Emerman, M. (2011). Human Trim5α has additional activities that are uncoupled from retroviral capsid recognition. Virology *409*, 113-120.

Tareen, S.U., Sawyer, S.L., Malik, H.S., and Emerman, M. (2009). An expanded clade of rodent Trim5 genes. Virology *385*, 473-483.

Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G.*, et al.* (2012). Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. Science.

Thielen, B.K., Klein, K.C., Walker, L.W., Rieck, M., Buckner, J.H., Tomblingson, G.W., and Lingappa, J.R. (2007). T cells contain an RNase-insensitive inhibitor of APOBEC3G deaminase activity. PLoS Pathog *3*, 1320-1334.

Tian, C., Yu, X., Zhang, W., Wang, T., Xu, R., and Yu, X.F. (2006). Differential requirement for conserved tryptophans in human immunodeficiency virus type 1 Vif for the selective suppression of APOBEC3G and APOBEC3F. J Virol *80*, 3112-3115.

Turelli, P., Vianin, S., and Trono, D. (2004). The innate antiretroviral factor APOBEC3G does not affect human LINE-1 retrotransposition in a cell culture assay. J Biol Chem *279*, 43371-43373.

Van Valen, L. (1973). A new evolutionary law. *Evolutionary Theory 1*, 1-30.

Verrelli, B.C., McDonald, J.H., Argyropoulos, G., Destro-Bisol, G., Froment, A., Drousiotou, A., Lefranc, G., Helal, A.N., Loiselet, J., and Tishkoff, S.A. (2002). Evidence for balancing selection from nucleotide sequence analyses of human G6PD. Am J Hum Genet *71*, 1112-1128.

Virgen, C.A., Kratovac, Z., Bieniasz, P.D., and Hatziioannou, T. (2008). Independent genesis of chimeric TRIM5-cyclophilin proteins in two primate species. Proc Natl Acad Sci U S A *105*, 3563-3568.

Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D., and Moran, J.V. (2001). Human L1 retrotransposition: cis preference versus trans complementation. Mol Cell Biol *21*, 1429-1439.

Wiegand, H.L., Doehle, B.P., Bogerd, H.P., and Cullen, B.R. (2004). A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. EMBO J *23*, 2451-2458.

Wilkins, C., and Gale, M. (2010). Recognition of viruses by cytoplasmic sensors. Curr Opin Immunol *22*, 41-47.

Wilson, S.J., Webb, B.L., Ylinen, L.M., Verschoor, E., Heeney, J.L., and Towers, G.J. (2008). Independent evolution of an antiviral TRIMCyp in rhesus macaques. Proc Natl Acad Sci U S A *105*, 3557-3562.

Wissing, S., Montano, M., Garcia-Perez, J.L., Moran, J.V., and Greene, W.C. (2011). Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. J Biol Chem *286*, 36427-36437.

Xu, H., Svarovskaia, E.S., Barr, R., Zhang, Y., Khan, M.A., Strebel, K., and Pathak, V.K. (2004). A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion. Proc Natl Acad Sci U S A *101*, 5652-5657.

Yamashita, M., and Emerman, M. (2004). Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. J Virol *78*, 5670-5678.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol *24*, 1586-1591.

Yang, Z., Wong, W.S., and Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol *22*, 1107-1118.

Yu, X., Yu, Y., Liu, B., Luo, K., Kong, W., Mao, P., and Yu, X.F. (2003). Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. Science *302*, 1056-1060.

Zhang, Q., Liu, Z., Mi, Z., Li, X., Jia, P., Zhou, J., Yin, X., You, X., Yu, L., Guo, F.*, et al.* (2011). High-throughput assay to identify inhibitors of Vpu-mediated down-regulation of cell surface BST-2. Antiviral Res *91*, 321-329.

Zheng, Y.H., Irwin, D., Kurosu, T., Tokunaga, K., Sata, T., and Peterlin, B.M. (2004). Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. J Virol *78*, 6073-6076.

CURRICULUM VITAE

Nisha K. Duggal

EDUCATION

University of Washington **Seattle, WA**
PhD, Molecular & Cellular Biology program *2012*

University of Michigan **Ann Arbor, MI**
B.S. with Distinction *2005*
Concentrations: English; Cell & Molecular Biology (High Honors)

FELLOWSHIPS and HONORS

APHL/CDC Emerging Infectious Disease Postdoctoral Fellowship *2012 – 2014*
Young Investigator Award, Conf. on Retroviruses & Opportunistic Infections *2012*
National Science Foundation Graduate Research Fellowship *2008 – 2011*
Conference Assistant Scholarship, Keystone Symposium *2011*
Global HIV Vaccine Enterprise grant *2010*
Best Grad Student Presentation award, West Coast Retrovirus Meeting *2009*
Dean's Merit Scholarship, University of Michigan *2001 – 2005*
University Honors, University of Michigan *2001 – 2005*
University of Michigan Program in Biology Scholarship *2004*
University of Michigan Summer Biomedical Research Fellowship *2004*

RESEARCH EXPERIENCE

Fred Hutchinson Cancer Research Center **Seattle, WA**
Graduate Research Assistant, laboratory of Dr. Michael Emerman *2007 - 2012*
Dissertation topic: The ancient evolution of Apobec3 antiviral activity in primates.

Fred Hutchinson Cancer Research Center **Seattle, WA**
Graduate Research Assistant, laboratory of Dr. Julie Overbaugh *Winter 2008*
Rotation topic: Role of immune activation in HIV-1 susceptibility in women.

University of Washington Department of Genome Sciences **Seattle, WA**
Graduate Research Assistant, laboratory of Dr. Joshua Akey *Spring 2008*
Rotation topic: Global population structure of *MYLK* in humans.

University of Michigan Medical School **Ann Arbor, MI**
Research Technician Associate, laboratory of Dr. Alice Telesnitsky *2005 – 2007*
Research topic: Determinants of homologous recombination of retroviruses.

University of Michigan Medical School **Ann Arbor, MI**
Undergraduate Research Assistant, laboratory of Dr. Jessica Schwartz *2003 – 2005*
Honors thesis topic: Growth hormone regulation of cellular transcription.

PUBLICATIONS

Duggal NK, Malik HS, Emerman M. Positive selection of Apobec3DE in chimpanzees has driven breadth in anti-viral activity. *J Virol*. 85(21):11361-71 (2011).

Duggal NK and Emerman M. Evolutionary conflicts between viruses and restriction factors shape immunity. *Nature Reviews Immunology*. doi:10.1038/nri3295 (2012).

Duggal NK and Emerman M. Loss-of-function polymorphisms in the Apobec3 locus in human populations**.** In preparation.

Duggal NK, Goo L, King SR, Telesnitsky A. Effects of identity minimization on Moloney murine leukemia virus template recognition, and frequent tertiary template-directed insertions during non-homologous recombination. *J Virol*. 81(22): 12156-68 (2007).

King SR, **Duggal NK,** Ndongmo CB, Pacut C, Telesnitsky A. Pseudodiploid genome organization aids full-length HIV-1 DNA synthesis. *J Virol*. 82(5): 2376-84 (2008).

Huo JS, McEachin RC, Cui TX, **Duggal NK,** Hai T, States DJ, Schwartz J. Profiles of growth hormone (GH)-regulated genes reveal time-dependent responses and identify a mechanism for regulation of activating transcription factor 3 by GH. *J Biol Chem* 281(7):4132-41 (2006).

PRESENTATIONS

Duggal NK, Emerman M, Malik HS. The diversity of Apobec3 antiviral activity in human populations. 19th Conference on Retroviruses and Opportunistic Infections, Seattle, WA, March 2012. Poster and oral presentation.

Duggal NK, Malik HS, Emerman M. Adaptive evolution of Apobec3DE provides evidence for a paleovirus infection in chimpanzee ancestors. HIV Evolution, Genetics, and Pathogenesis Keystone Symposia, Whistler, B.C., March 2011. Poster presentation.

Duggal NK, Malik HS, Emerman M. Evidence for a paleovirus infection in chimpanzee ancestors. Retroviruses Meeting at Cold Spring Harbor Laboratory, Long Island, NY, May 2010. Oral presentation.

Duggal NK, Malik HS, Emerman M. Apobec3DE: The fastest evolving Apobec3 gene. The 16th West Coast Retrovirus meeting, Palm Springs, CA, Oct. 2009. Oral presentation.

Duggal NK, Goo L, King SR, Telesnitsky A. Minimal homology required for MoMLV recombination and outcomes of forced non-homologous template switch. Retroviruses Meeting at Cold Spring Harbor Laboratory, Long Island, NY, May 2007. Oral presentation.

TEACHING EXPERIENCE

University of Washington Teaching Assistant	**Seattle, WA**
Science Education Partnership	*2009, 2011*
University of Washington Teaching Assistant	**Seattle, WA**
Introduction to Cell Biology	*2009*
University of Michigan Undergraduate Teaching Assistant	**Ann Arbor, MI**
Introductory Biochemistry	*2003*

LEADERSHIP and ACTIVITIES

University of Washington Undergraduate Research program	**Seattle, WA**
Mentorto undergraduate students	*2008 – 2011*
Molecular and Cellular Biology graduate student meetings	**Seattle, WA**
Co-organizer	*2008 – 2011*
Global HIV Vaccine Enterprise town hall meeting	**Seattle, WA**
Meeting organizer, "Let's Talk HIV Prevention"	*2010*
Michigan Infectious Disease International Scholars	**Ann Arbor, MI**
Program assistant	*2005 – 2007*
World AIDS Day coalition	**Ann Arbor, MI**
Planning committee member	*2005*
Student Global AIDS Campaign	**Ann Arbor, MI**
Treasurer	*2004 – 2005*