The Evolution of Endogenous Viral Elements

Edward C. Holmes[1,2,*]
[1]Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA
[2]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA
*Correspondence: ech15@psu.edu
DOI 10.1016/j.chom.2011.09.002

Endogenous retroviruses are a common component of the eukaryotic genome, and their evolution and potential function have attracted considerable interest. More surprising was the recent discovery that eukaryotic genomes contain sequences from RNA viruses that have no DNA stage in their life cycle. Similarly, several single-stranded DNA viruses have left integrated copies in their host genomes. This review explores some major evolutionary aspects arising from the discovery of these endogenous viral elements (EVEs). In particular, the reasons for the bias toward EVEs derived from negative-sense RNA viruses are considered, as well as what they tell us about the long-term "arms races" between hosts and viruses, characterized by episodes of selection and counter-selection. Most dramatically, the presence of orthologous EVEs in divergent hosts demonstrates that some viral families have ancestries dating back almost 100 million years, and hence are far older than expected from the phylogenetic analysis of their exogenous relatives.

Although it has long been known that the eukaryote genome contains a myriad of complete and partial relatives of retroviruses called endogenous retroviruses that are now inherited passively with the host genetic component, it was surprising to discover that sequences of RNA viruses, which do not make a DNA intermediate and usually do not enter the host cell nucleus, were also present in eukaryotic genomes (Crochu et al., 2004). Although integrated copies of single-strand DNA (ssDNA) viruses were found in the genomes of plant viruses some years ago (Bejarano et al., 1996), a number of endogenized ssDNA viruses were recently described in a diverse set of animal genomes (Belyi et al., 2010a; Katzourakis and Gifford, 2010). Such an array of virally derived genetic material raises a number of important evolutionary questions, which forms the basis of this review. Perhaps the most important theme is that more than expanding our basic knowledge of the composition of eukaryotic genomes, the presence of endogenous viruses has had a profound impact on our understanding of the time-scale of virus evolution, the consequences of which are yet to be fully understood.

The Types and Phylogenetic Distribution of Endogenous Viruses

Although endogenous viruses have been classified in different ways, a useful collective term for them all that reflects their generally fragmentary nature is "endogenous viral elements" (EVEs) (Katzourakis and Gifford, 2010). From here on, the term EVE will be used to refer to all endogenous viruses, whether derived from retroviruses, DNA viruses, or RNA viruses, although this review will generally focus on the latter group. EVEs are generated when a double-stranded DNA (dsDNA) copy of the viral genome is integrated into the host germline (Figure 1). Although the production of dsDNA intermediates that integrate into host genomic DNA is an obligatory part of the retroviral life cycle, germline as opposed to somatic cell integration is expected to be a rare event. Despite this, endogenous retroviruses are surprisingly common; for example, approximately 5%–8% of the human genome is composed of endogenous retroviruses (Katzourakis and Tristem, 2005). These comprise at least 31 distinct families, such that there have been at least 31 separate integration events, and likely many more. In some animal species, it has even proven possible to see retroviral endogenization in action (Tarlinton et al., 2006). Not surprisingly, those EVEs derived from nonretroviruses are far rarer and hence represent the consequence of sporadic evolutionary events. In addition, while effectively complete genomes of endogenous retroviruses are relatively commonplace, this is not the case for those EVEs generated by other types of virus, which are usually composed of partial genomic fragments.

Although the presence of retroviral EVEs was described many years ago (Benveniste and Todaro, 1974; Weiss et al., 1973), the first description of a gene sequence of an RNA virus (that replicate using an RNA-dependent RNA polymerase [RdRp]) integrated into the host genome did not occur until 2004. This example involved the insect flavivirus cell fusing agent (CFA), fragments of which were found to be integrated into the genomes of *Aedes* spp. mosquitoes (Crochu et al., 2004), and which has also been observed in some other insect flaviviruses (Roiz et al., 2009). Shortly afterward, integrated sequences of the RNA virus Potato virus Y (PVY, a potyvirus) were found in the genomes of some grapevine varieties (Tanne and Sela, 2005). Since this time, a number of other endogenous RNA viruses have been discovered, comprising bornaviruses (also referred to as endogenous Borna-like N elements [EBLN]; Figure 2) (Belyi et al., 2010b; Horie et al., 2010), bunyaviruses (Katzourakis and Gifford, 2010), filoviruses (Belyi et al., 2010b; Taylor et al., 2010), orthomyxoviruses (Katzourakis and Gifford, 2010), reoviruses (Katzourakis and Gifford 2010), and rhabdoviruses (Katzourakis and Gifford, 2010), although usually at very low copy numbers (i.e., less than 100 elements per genome). A list of animal EVEs is provided in Table 1, with bornaviruses being the only endogenous RNA viruses found in the human genome. Similarly, endogenous viruses have been observed in various fungal (Frank and Wolfe, 2009; Taylor and Bruenn, 2009) and bacterial (Salanoubat et al., 2002) genomes.
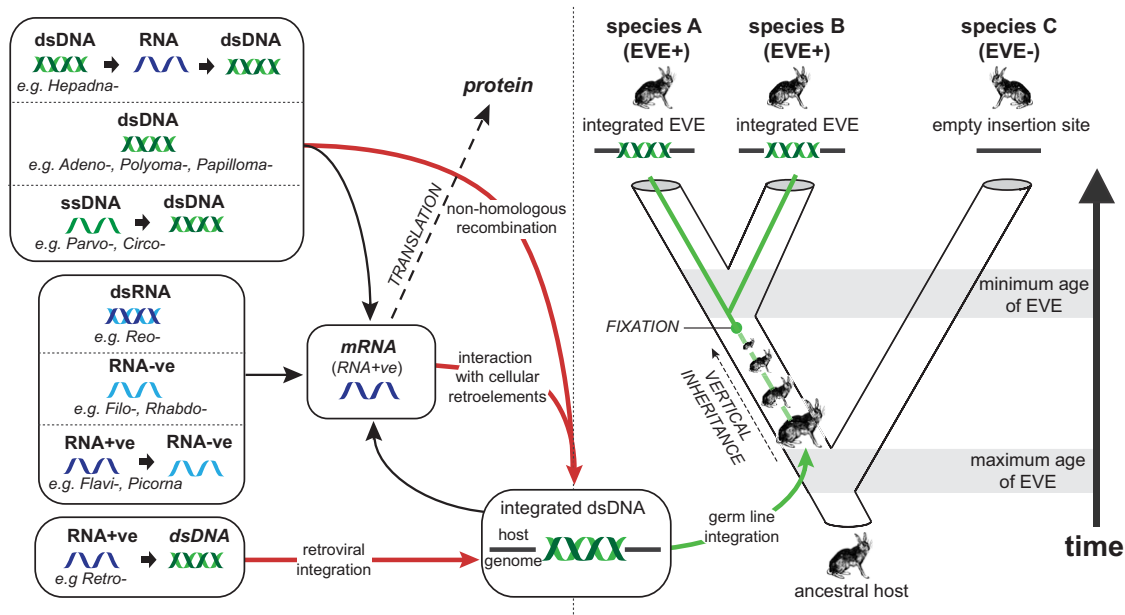
Figure 1. Processes Involved in the Generation of EVEs, and How EVEs Can Be Used to Estimate the Age of Viruses, Using Both RNA and DNA Viruses as Examples
The presence of EVEs in related species A and B and integrated into the same genomic position such that they are orthologous indicates that this integration event occurred prior to the divergence of these two species (species C is an outgroup). If it is known when species A and B diverged, then the minimum age of the insertion event can also be estimated. See Katzourakis and Gifford (2010) for more details. Figure kindly provided by Rob Gifford.

As well as the expanded catalog of endogenous RNA viruses, the sequences of a number of small DNA viruses are also integrated into host genomes, namely the circovirus (Belyi et al., 2010a), geminivirus (Bejarano et al., 1996), and parvovirus (Belyi et al., 2010a; Kapoor et al., 2010) families of ssDNA viruses, as well as the hepadnaviruses (dsDNA) (Gilbert and Feschotte, 2010). For example, endogenous fragments of the *Parvoviridae* are found in mammalian hosts as diverse as cats, elephants, platypus, and wallabies. With the continual increase in the number of eukaryotic genome sequences, it is certain that more EVEs will be described. In addition, it is likely that the phylogenetic signal for the integration of some very ancient EVEs, particularly those that occurred early in eukaryote evolution and that evolve without functional constraints, has been lost through the accumulation of multiple nucleotide substitutions.

Of those EVEs recently discovered, perhaps the most surprising were the endogenous filoviruses (Figure 3). Although novel exogenous filoviruses have recently been discovered (Barrette et al., 2011), they generally infect a small number of mammalian species—particularly primate and bat species from equatorial Africa—with occasional high-profile spillovers into human populations. Indeed, there has been a long running debate as to the natural reservoir of filoviruses, with various species of fruit bat being perhaps the most likely candidate (Leroy et al., 2005). In contrast, endogenous filoviruses have been detected in the genomes of diverse mammalian taxa, including both placental and marsupial mammals (Belyi et al., 2010b; Taylor et al., 2010). While much of the evolution of these endogenous filoviruses can be measured on time scales of millions of years—especially as EVEs from rat and mouse are inserted at homologous loci, strongly suggesting that they

have codiverged with these species—it is evident that a number of independent insertion events have occurred (Taylor et al., 2010). Most notably, those filovirus EVEs infecting placental and marsupial mammals are not sister taxa, as expected given ancient virus-host codivergence, with the EVEs from marsupials more closely related to the exogenous filoviruses. This is a remarkable observation, given the geographic separation of these two groups; most known exogenous filoviruses are of African origin, whereas the marsupial endogenous filoviruses are largely from Australian species, with a single representative found in an American opossum that constitutes the closest relative of the exogenous filoviruses (Taylor et al., 2010). Such a disjunct distribution is highly suggestive of the presence of further, currently uncharacterized, exogenous filoviruses.

Despite the evident under- and biased sampling of EVEs, there is a striking imbalance in the taxonomic origins of those EVEs derived from RNA viruses described to date; at the time of writing, only eight families of RNA viruses have been shown to possess endogenous relatives, five of which represent viruses with negative-sense genomes (ssRNA⁻ viruses), and three of these falling into the *Mononegavirales*—a higher-order grouping of multiple families of ssRNA⁻ viruses with unsegmented genomes (Table 1). In addition, those EVEs derived from positive-sense RNA viruses (ssRNA⁺ viruses) are at extremely low copy number; one genomic copy in the case of the *Reoviridae*, five in the case of the *Flaviviridae*, and probably a small number in PVY. Such a bias toward ssRNA⁻ viruses merits explanation. In the case of the bornavirus EVEs, at least part of the explanation must relate to their nuclear replication cycle, increasing the chances of endogenization (see below). Similarly, there will be an elevated chance of endogenization for those viruses that
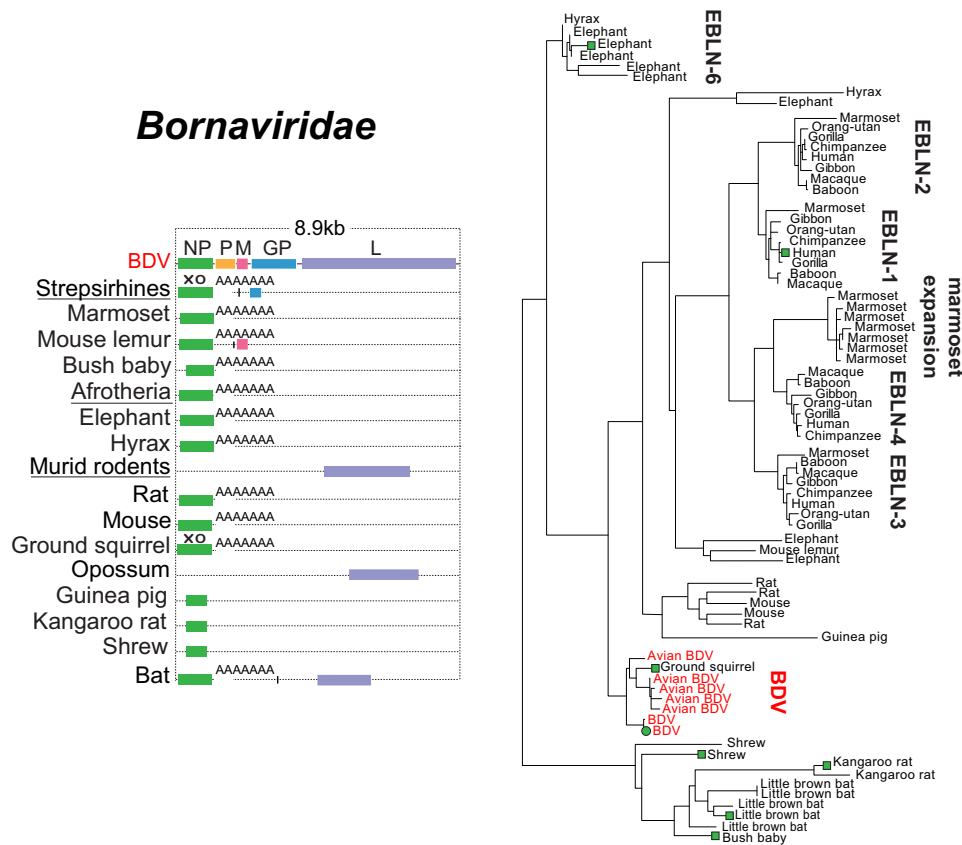
Figure 2. Genomic Structures and Phylogenetic Distribution of Exogenous and Endogenous Bornaviruses, Members of the Order *Mononegavirales*

BDV, exogenous Borna disease virus (shown in red); EBLN, endogenous viruses. For each species, the most intact endogenous elements are shown relative to a representative complete exogenous virus genome. Squares of a specific color on the phylogeny indicate EVEs and the viral gene they represent. Note the bias toward integrated NP sequences. EVEs with poly-A tails are shown, while intact ORFs are marked by an "O" symbol and expressed genes by an "X" symbol. Vertical lines identify noncontiguous genes. Adapted from Katzourakis and Gifford (2010), which should be consulted for more details. Original figure kindly provided by Rob Gifford.

cause persistent as opposed to acute host infections. However, as ssRNA⁺ viruses are as likely to generate persistent infections as ssRNA⁻ viruses, this cannot explain the extreme distribution bias. It is also possible that some EVEs are better able to achieve germline integration than others (Horie et al., 2010; Johnson 2010), although the determinants of this process are currently unclear, nor is why it might occur more frequently with ssRNA⁻ viruses. One possibility is that the messenger RNAs (mRNAs) of ssRNA⁻ viruses are more favorable templates for L1-mediated reverse transcription than those of other viruses (Horie et al., 2010), although why is again unclear.

One clue to the preponderance of ssRNA⁻ viruses comes from the strong bias toward the integration of nucleoprotein (*NP*) genes (Figures 2 and 3). The *Mononegavirales* share a common genome organization, in which the *NP* gene has the most 3′ location and the *L* gene, which encodes the RdRp, the most 5′. In all *Mononegavirales* except the filoviruses genes are transcribed in a sequential manner from 3′ to 5′ and with stepwise attenuation. This results in discrete mRNAs for each gene and means that the most 3′ gene (*NP*) is the most abundant RNA, and the most 5′ (*L*) gene the least abundant, and hence that whole genomes of *Mononegavirales* are not expected to be endogenized. That

most EVEs represent *NP* genes also suggests that endogenization is at least in part a function of relative mRNA abundance. Indeed, many ssRNA⁺ viruses produce a single large polyprotein, and it may be that L1-mediated reverse transcription occurs more efficiently on the shorter mRNAs produced by ssRNA⁻ viruses. Similarly, it is possible that the life cycle of ssRNA⁺ viruses, in which translation occurs before transcription, also influences mRNA abundance and hence the probability of endogenization.

Generation and Fitness of Endogenous Viruses
As noted above, one reason why the existence of endogenous RNA viruses came as a surprise to researchers is that they require two unusual steps in the viral life cycle: first, the viral genetic material needs to enter the cell nucleus, when the fact that RNA viruses carry their own RdRp means that they usually (with a few exceptions) only inhabit the cytoplasm, and second, ssRNA needs to be converted into dsDNA (Figure 1). There are currently little meaningful data to determine how the first process occurs, although it is striking that endogenous bornaviruses are particularly commonplace and these are one of the few families of RNA viruses that replicate within the cell nucleus (Horie et al., 2010).

Table 1. Distribution and Age of Endogenous Viral Elements Derived from RNA and DNA Viruses Found in Animal Genomes

Viral Family and Type	Host Range	Number of Elements	Estimated Minimum Age	Reference
Bornaviridae (ssRNA⁻)[a]	Mammals[b]	67	93 MYA	Belyi et al. (2010b), Horie et al. (2010), Katzourakis and Gifford (2010)
Filoviridae (ssRNA⁻)[a]	Mammals	25	30 MYA[c]	Belyi et al. (2010b), Katzourakis and Gifford (2010), Taylor et al. (2010)
Bunyaviridae (ssRNA⁻)[a]	Insects	40	Unknown	Katzourakis and Gifford (2010)
Rhabdoviridae (ssRNA⁻)[a]	Insects	143	Unknown	Katzourakis and Gifford (2010)
Orthomyxoviridae (ssRNA⁻)	Insects	1	Unknown	Katzourakis and Gifford (2010)
Reoviridae (dsRNA)	Insects	1	Unknown	Katzourakis and Gifford (2010)
Flaviviridae (ssRNA⁺)	Insects	5	Unknown	Crochu et al. (2004), Katzourakis and Gifford (2010)
Parvoviridae (ssDNA)	Mammals	99	30 MYA[c]	Belyi et al. (2010a), Katzourakis and Gifford (2010)
Circoviridae (ssDNA)	Mammals	5	68 MYA	Belyi et al. (2010a), Katzourakis and Gifford (2010)
Hepadnaviridae (dsDNA)	Birds	8	>19 MYA	Gilbert and Feschotte, 2010, Katzourakis and Gifford (2010)

MYA, million years ago.
[a] *Mononegavirales*.
[b] Includes humans.
[c] Will vary according to the date used for the divergence of rat and mouse.

There are, however, a number of plausible ways in which the conversion from ssRNA to dsDNA can occur. Perhaps the most likely involves a reverse transcription step using the reverse transcriptase (RT) present in the cellular retroelements that are abundant in eukaryotic genomes. For example, long interspersed nucleotide elements (LINES) are a common component of vertebrate genomes, particularly members of the L1 family, and therefore are a potentially rich source of RT. In fact, the flanking sequences of some EBLNs possess signatures suggestive of L1-mediated reverse transcription, such as the presence of 3′ poly-A tails and target site duplications (Belyi et al., 2010b; Horie et al., 2010), while the endogenous PVY elements in plants possess direct repeats and lie within a sequence itself flanked by inverted repeats (Tanne and Sela, 2005), compatible with transposable element-mediated integration. More directly, nonhomologous recombination between an exogenous RNA virus and an intracisternal A-type particle (IAP) retrotransposon has been observed to result in the reverse transcription and cellular integration of viral RNAs (Geuking et al., 2009). Alternatively, it is possible that the RT is provided by an exogenous retrovirus that is infecting the host at the same time, and particularly where this exogenous infection is associated with a high viral copy number and abundant RT. The situation is rather different for the endogenized DNA viruses. As small ssDNA

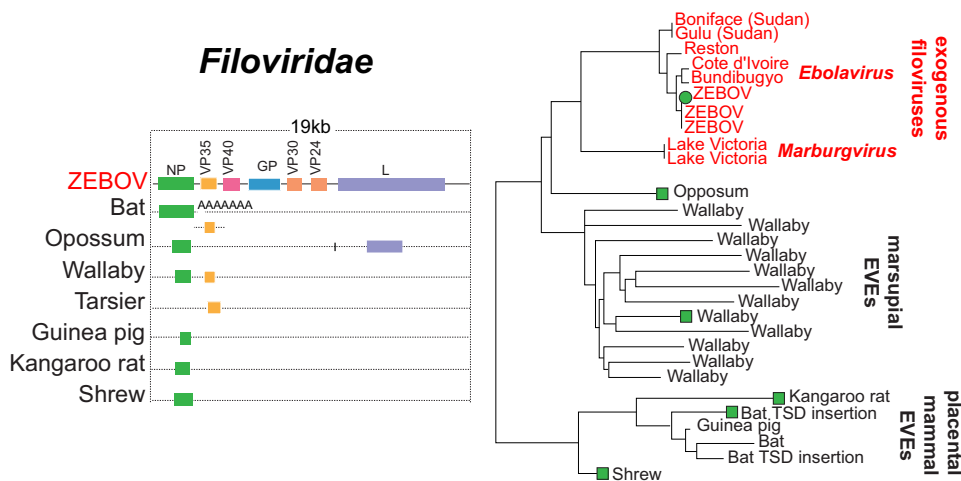**Figure 3. Genomic Structures and Phylogenetic Distribution of Exogenous and Endogenous Filoviruses, Members of the Viral Order *Mononegavirales***
ZEBOV, exogenous Ebola virus Zaire (shown in red); TSD, target site duplication. Other labeling is the same as that in Figure 2. Adapted from Katzourakis and Gifford (2010), which should be consulted for more details. Original figure kindly provided by Rob Gifford.
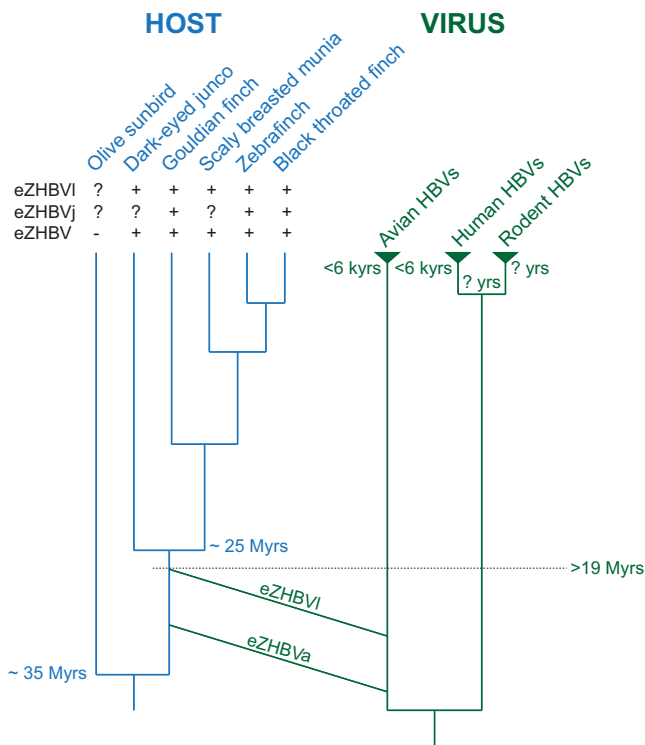
HOST   **VIRUS**

Olive sunbird
Dark-eyed junco
Gouldian finch
Scaly breasted munia
Zebrafinch
Black throated finch

Avian HBVs   Human HBVs   Rodent HBVs

eZHBVl	?	+	+	?	+	+
eZHBVj	?	?	+	?	+	+
eZHBV	–	+	+	+	+	+

<6 kyrs   <6 kyrs   ? yrs   ? yrs

~ 25 Myrs

>19 Myrs

eZHBVl

eZHBVa

~ 35 Myrs

Figure 4. Evolution of Exogenous and Endogenous Avian Hepadnaviruses

The phylogeny of the bird host species is shown in blue on the left of the figure, while that of the hepadnaviruses (both avian and mammalian) is shown in green on the right. Estimates of divergence time are also shown in both cases; note the huge difference between the time scale of the host (millions of years) and virus (thousands of years) trees. The presence or absence of orthologous eZHBV insertions in various bird species (i.e., eZHBVl, eZHBVj, and eZHBVa) are denoted by the "+" and "−" symbols, respectively, while a question mark means uncertain status. The likely phylogenetic placement of the germline integrations producing eZHBVl and eZHBVa are shown by the green lines connecting the host and virus trees. Using estimated bird host divergence times of 25 MYA and 35 MYA, it is conservatively estimated that the integration of eZHBVa and eZHBVl must have occurred at least 19 MYA, as shown by the dashed line. Adapted from Gilbert and Feschotte (2010), which should be consulted for more details. Original figure kindly provided by Clément Gilbert.

viruses such as parvoviruses utilize host DNA polymerases for their replication, and therefore enter the cell nucleus, their endogenization should not come as a surprise. A similar story can be told for the endogenous hepadnaviruses, the first of which was discovered in the zebra finch and designated eZHBV (Gilbert and Feschotte, 2010) (Figure 4). Importantly, hepadnaviruses possess dsDNA genomes, utilize RT, replicate in the nucleus, and integrated copies are seen with the human form of the virus where they are associated with liver cancer (Bonilla Guerrero and Roberts, 2005). Although the precise mechanisms of genomic integration are unknown (Gilbert and Feschotte, 2010), hepadnaviruses clearly possess a number of the necessary attributes for endogenization. In this context, it is interesting that endogenous copies of another important agent of human cancer—the papillomaviruses—have yet to be discovered, even though they possess dsDNA genomes, enter the cell nucleus, and are commonplace in vertebrates (Bernard et al., 1994).

While there is mounting evidence for the presence of virus genetic material in host genomes, there are few examples of

host genetic material integrated into the genomes of RNA and small DNA viruses, such that lateral gene transfer is largely one-way traffic. This is most likely a reflection of the extreme size constraints faced by RNA and small DNA viruses, which rarely have genomes >20,000 nt and are characterized by a lack of truly nonfunctional genomic regions. Importantly, not only do RNA and ssDNA viruses both have very small genome sizes, but they also have a relatively high mutation rate per nucleotide (Sanjuán et al., 2010), such that mutational load likely limits genome size (Holmes, 2011). Accordingly, an increase in viral genome size that would follow the insertion of a host gene would result in a concomitant increase in mutational load and hence a reduction in viral fitness (Holmes, 2009). One of the few cases of RNA virus capture and maintenance of a host gene concerns the ExoN domain of coronaviruses and roniviruses, which encodes a 3′-to-5′ exoribonuclease. This domain shares a distant similarity with host cellular proteins of the DEDD superfamily of exonucleases that are involved in proofreading (Minskaia et al., 2006; Snijder et al., 2003), and a proofreading mechanism has recently been demonstrated in coronaviruses (Denison et al., 2011). Hence, coronaviruses and roniviruses may be able to reduce, to some extent, the high error rate normally associated with RdRp replication, in turn allowing an increase in genome size up to ~30,000 nt and making them the largest of all RNA viruses. In contrast, eukaryotes generally experience weaker constraints on genome size and are able to carry extra genetic material if this represents a minor fitness cost, as is often likely to be the case.

The EVEs we know today must only be a small subset of those that have existed in the past; many others will have been lost by the chance process of genetic drift, which is the fate of most mutations at low frequency, even those that are selectively advantageous. Indeed, the extra cost in resources required for the replication of a single EVE will be minimal, and perhaps effectively neutral if they fall into a genomic region that contains few genes. The potential selective neutrality of EVEs is especially likely in mammalian species that are characterized by small effective population sizes (N_e), such that genetic drift will dominate evolutionary dynamics (Ohta 1992) and perhaps play a major role in shaping genomic architectures (Lynch and Conery, 2003).

Other EVEs may have been removed by purifying selection because they reduce organismal fitness. In particular, human endogenous retroviruses are usually located in genomic regions away from genes, whereas the integration sites of (presumably recent) exogenous retroviruses are often close to genes, suggesting that there is a selective cost in having EVEs located too close to genic regions (Medstrand et al., 2002). However, it is important to note that the negative selection pressure exerted by endogenous viruses on their hosts is likely to be tiny compared to the possible selective costs due to exogenous viruses, which could induce mass host mortality.

While most EVEs are likely to be functionally defective, either because they were inserted as partial proteins or have more recently accumulated stop codon mutations, some may have retained or acquired a function, and a small number appear to be expressed as mRNA (Katzourakis and Gifford, 2010) (Figure 2). In these cases EVEs could be selectively advantageous to the host. The most obvious selective benefit of EVEs

is that they confer protection against related exogenous viruses, perhaps by triggering some aspect of the host innate immune response or, in the case of functionally intact EVE proteins, through the expression of proteins that act as immunogens. A good example is provided by the endogenous and exogenous versions of Jaagsiekte sheep retrovirus (Arnaud et al., 2007). Protection here stems from the similar Gag proteins exhibited by the endogenous and exogenous forms of these viruses. The Gag protein from the endogenous virus interacts and coassembles with the exogenous Gag proteins, making chimeric and defective viral particles (Murcia et al., 2007). Notably, some exogenous viruses have escaped from this restriction, indicative of a virus-host evolutionary "arms race" (Arnaud et al., 2007), and these are discussed in more detail below. However, immunity is not the only possible functional benefit for the host. Some wasp species encode genomic relatives of nudiviruses that produce virus-like particles that allow them to parasitize the larvae of lepidopterans by manipulating host responses (Bézier et al., 2009), EVEs integrated near genes may be co-opted as promoters or *cis*-regulatory elements (Feschotte, 2008), and the protein syncytin that is involved in the development of the human placenta is derived from the envelope protein of an endogenous retrovirus (HERV-W) (Mi et al., 2000).

Another line of evidence suggestive of EVE function is that their sequences are sometimes more conserved than might be expected from an entirely neutral evolutionary process. A case in point are some endogenous bornaviruses that have acquired fewer stop codon mutations than expected given their antiquity, such that they are subject to the process of purifying selection indicative of functional constraint (Belyi et al., 2010b; Katzourakis and Gifford, 2010). However, the precise function of these bornavirus EVEs is unknown, and that endogenous RNA viruses are so rarely found in host genomes, are fragmentary, and often contain debilitating mutations suggests that they are usually chance passengers in our genome.

EVEs and the Time Scale of Viral Evolution

Although many of the possible evolutionary consequences of EVEs merit more-detailed investigation, their discovery has already contributed to a radical shift in our understanding of the time scale of virus evolution. As pointed about by a number of authors, the power of EVEs in this context is that they effectively represent a "fossil record" of past viral infections, albeit a very biased one (Emerman and Malik, 2010). The key point here is that once integrated into host genomes, EVEs cease to evolve with the very high substitution rates that characterize exogenous RNA and small DNA viruses (Holmes, 2009) and instead replicate using high-fidelity host DNA polymerases and probably experience fewer replications per unit time. This will result in a dramatic reduction in evolutionary rate, from the virus scale (usually around 10^{-3} nucleotide substitutions per site, per year) (Duffy et al., 2008) to the host scale ($\sim 10^{-9}$ subs/site/year). If the mutational differences between endogenous viruses are known to occur after integration, such as those between duplicated EVEs, those observed between the LTRs of retroviral EVEs, or, most powerfully of all, when there is clear evidence for virus-host codivergence—such as when EVEs are integrated into similar genomic positions in sister taxa, which definitively shows that they are orthologous—then an evolutionary time

scale can be estimated in a relatively straightforward manner using host divergence times as calibration points (Figures 1 and 4). However, as with all fossils, there is uncertainty about the exact timing of the evolutionary strata, particularly in cases where it is impossible to exclude the independent integration into different species. Indeed, as cross-species transmission between related hosts appears to be a major mode of exogenous RNA virus evolution (Kitchen et al., 2011), independent insertion may sometimes be difficult to exclude unless the EVEs in question occupy orthologous loci.

Those studies of the time scale of EVE evolution undertaken to date have revealed that some virus families are of great antiquity. Indeed, the analysis of EVEs has generally painted a radically different picture of the time scale of viral evolution than that inferred from molecular clock studies using "heterochronous" samples from a single virus; that is, molecular clock studies of the differences in tree branch lengths between exogenous viruses sampled at different time points during epidemiological history (Drummond et al., 2003). The best described case is that of the primate lentiviruses, a group of pathogens that includes the human (HIV) and simian (SIV) immunodeficiency viruses. Estimates of the age of primate lentiviruses based on the use of heterochronous sequences result in time scales of thousands (Sharp et al., 2001) or even hundreds (Wertheim and Worobey, 2009) of years, which are surprisingly recent given the great genetic and geographic diversity of these viruses in nonhuman primates from Africa (Hahn et al., 2000). In contrast, the presence of endogenous lentiviruses in lemurs indicates that these viruses have circulated in some primates for at least several million years (Gifford et al., 2008), while biogeographic analyses of SIV are suggestive of time scales of hundreds of thousands of years and perhaps far longer (Worobey et al., 2010). In addition, the presence of RELIK (rabbit endogenous lentivirus type K) (Katzourakis et al., 2007) elements in related lagomorph species (rabbits and hares) suggests that lentiviruses have been present in other mammalian orders for at least 12 million years (Keckesova et al., 2009; van der Loo et al., 2009).

Similar stories can be told for other viruses. The observation of endogenous hepadnaviruses integrated at the same genomic positions in bird species provides compelling evidence that hepadnaviruses have been in circulating in birds for at least 19 million years and perhaps as long as 40 million years (Figure 4) (Gilbert and Feschotte, 2010). Such ancient evolution sits in stark contrast to molecular clock studies of hepadnavirus evolution based on the use of contemporary avian viruses, in which evolutionary history has been measured in time scales of thousands of years (Zhou and Holmes, 2007). Similarly, the earliest integrations of bornaviruses are proposed to have occurred almost 100 million years ago (MYA) (Katzourakis and Gifford, 2010) and those of circoviruses at over 50 MYA (Belyi et al., 2010a; Katzourakis and Gifford, 2010) (Table 1). Not only do these data describe a far older history for some viral families than previously anticipated, but such ancient virus-host codivergence sits in marked contrast to the process of recent cross-species transmission and emergence (i.e., host jumping) that characterizes many exogenous viruses. A good example is provided by the parvoviruses. While some endogenous parvoviruses have been associated with their hosts for perhaps 30 million years (Kapoor et al., 2010; Katzourakis and Gifford, 2010), these small

DNA viruses represent a textbook example of recent viral emergence. In particular, it is well known that feline panleukopenia virus (FPV) jumped to dogs in the late 1970s and resulted in the pandemic spread of canine parvovirus (CPV) (Parrish 1990) and has evolved rapidly since this time (Shackelton et al., 2005). Finally, early divergences also have been proposed for the families of human endogenous retroviruses (HERVs); the origin of the HERV-A family has been proposed at between 57 and 92 MYA, that of HERV-K to between 51 and 83 MYA, and HERV-L at between 62 and 100 MYA (Katzourakis and Tristem, 2005).

The presence of EVEs has unequivocally shown that some viral families are far older than inferred from the analysis of contemporary virus sequences. Although it has always been likely that viruses are ancient, and there is growing evidence that viruses are descendants of replicating elements that existed during a precellular world (Holmes, 2011; Koonin et al., 2006), it is perhaps surprising that some extant families of RNA virus arose so very long ago. In particular, as most exogenous RNA viruses evolve extremely rapidly, those viruses that existed millions of years ago and fossilized as EVEs would be expected to be unrecognizably divergent from their contemporary relatives, such that no meaningful multiple sequence alignment, let alone phylogenetic analysis, could be undertaken (Holmes, 2003a). However, that this is demonstrably not the case with the EVEs obtained to date (while acknowledging that more divergent EVEs might not be detected) raises the question of how ancient viruses can be so relatively well conserved in sequence. Of course, it is this sequence conservation that makes most molecular clock analyses give very recent dates of viral origin. Reconciling these profoundly different time scales for viral evolution—generally recent from the analysis of contemporary viral sequences, and usually ancient from the analysis of endogenous viruses—is one of the most pressing questions in studies of virus evolution and is likely to require new models of nucleotide and amino acid substitution that recognize the intricacies of viral evolution (Holmes, 2003a; Wertheim and Kosakovsky Pond, 2011).

Inference of the time scale of EVE evolution may also shed light on some other interesting problems in viral evolution and epidemiology. For example, the match between host and virus phylogenies suggests that hantaviruses have codiverged with rodent species for many millions of years (Plyusnin and Morzunov, 2001). However, more recent analyses have shown that hantaviruses jump species boundaries more frequently than previously realized and possess high rates of nucleotide substitution in the short term (Ramsden et al., 2008; Ramsden et al., 2009), which suggests a far more evolutionary history. Clearly, the discovery of endogenous hantaviruses (should they exist), and particularly those where the time scale seems to match that of the rodent hosts, would go a long way to proving that hantaviruses really do have ancient ancestries.

EVEs may also provide some useful insights into viral macroevolution. New virus lineages are generated when viruses diverge within a single host species, for example by adapting to new cell types or following the cross-species transmission to new host species. Phylogenetic studies suggest that cross-species transmission may be the most common mechanism underlying viral speciation, although often involving jumps among closely related host species (Kitchen et al., 2011).

Lineage death occurs when viruses are unable to find a sufficient number of susceptible hosts to sustain their transmission, if the infected host population suffers extinction, or because they are outcompeted by other viruses in the population. As an example of lineage death, simian foamy viruses (SFVs) are commonly found as exogenous agents in anthropoid primates, including great apes, and have seemingly codiverged with these species for more than 30 million years (Liu et al., 2008; Switzer et al., 2005), while the presence of endogenous copies suggests a far older association with placental mammals (Katzourakis et al., 2009). However, foamy viruses only appear in human populations as the result of transient spillovers from other primates. This suggests that SFV was lost from human populations during our early evolution, perhaps due to the small effective size of early human populations, which would increase the likelihood of stochastic extinction, or through the absence of the biting behavior that is central to viral transmission. Although detecting the extinction of viral lineages is a difficult task, with few genomic signatures, the presence of endogenous viruses represents the best evidence that a specific viral lineage has existed in the past (Katzourakis et al., 2009).

The Host-Virus Arms Race

Additional evidence for the antiquity of some viral families, albeit less direct, comes from observations that some host genes have been subject to selection pressure from viruses for millions of years. Because viral infections may impose a major fitness cost on their hosts, there will also be a strong selection pressure for hosts to evolve an effective antiviral response; alleles in antiviral genes that are able to prevent or clear viral infections will have a major selective benefit and likely spread rapidly through a host population. Evidence for the strength of this selective process is the remarkable number of genes that are associated with controlling pathogen infections. There are, for example, over 220 gene loci in the human major histocompatibility complex (MHC) (The MHC Sequencing Consortium 1999), which exhibit considerable allelic variation, and positive selection is routinely detected in genes involved in immune responses (Yang and Bielawski, 2000). More generally, the strongly deleterious consequences of parasite infections on hosts have been touted as a possible explanation for the long-term maintenance of sexual reproduction in eukaryotes (Hamilton et al., 1990). This selective process in the host will, in turn, result in a strong selection pressure for the virus to evade these host immune responses, giving rise to an evolutionary arms race between host and virus (Meyerson and Sawyer, 2011; Sawyer et al., 2004).

That the counter selection pressures on host and virus are very strong makes it possible to use the genomic signature of positive selection as a way of detecting those host genes, or specific regions within host genes, that are associated with strong antiviral infections in the past; specifically, a high ratio of nonsynonymous to synonymous nucleotide substitutions per site (i.e., a d_N/d_S ratio > 1) is a tell tale sign of past adaptive evolution (Yang and Bielawski, 2000), although care must be taken when making inferences in this area as false-positive results are a regular occurrence (Nozawa et al., 2009). In addition, if these bouts of positive selection occur on specific branches of the host phylogeny and the time scale of host evolution is known,

then it is possible to place these virus-host arms races in real time. The use of just such an approach has proven it possible to extend the likely age of some viral families to many millions of years (Emerman and Malik, 2010).

Research over the last decade has documented a number of host "restriction factors" that recognize viruses and inhibit their replication, which are involved in arms races with a variety of viruses. These restriction factors are particularly well described in primates, with the APOBEC (apolipoprotein B editing catalytic polypeptide) (Mangeat et al., 2003), BST-2/tetherin (Evans et al., 2010), and TRIM (tri-partite motif) (Nisole et al., 2005) protein families prominent among them (Meyerson and Sawyer, 2011). The product of the *TRIM5* gene—TRIM5α—is a restriction factor directed against the capsid protein of retroviruses. Notably, human TRIM5α has an elevated d_N/d_S on the branch after its divergence from chimpanzees, indicative of virally induced adaptive evolution in the last 6 million years or so. While this evolutionary pattern is indicative of a past retroviral induced selection, identifying the exact causative virus(es) is difficult, as will be the case in all studies of arms races based on analyses of host genomic data. Single mutations in both TRIM5α and viral capsids can lead to important differences in the specificity of viral recognition, highlighting the intricacy of this arms race (and single, selectively advantageous, amino acid changes are usually very difficult to detect with available bioinformatic methods). As TRIM5 is able to block retroviral infection of new primate hosts, it may also play a central role in cross-species transmission and emergence (Kirmaier et al., 2010; Stremlau et al., 2004).

Similar virus versus host restriction factor arms races have been documented in *APOBEC3G* and the antiviral protein kinase R (*PKR*) gene, the latter having a complex evolutionary interaction with poxviruses (Elde et al., 2009). *APOBEC3G,* along with its antiretroviral relative *APOBEC3F*, is a member of a gene family involved in the editing of RNA and/or DNA through the deamination of cytosine. When directed to the reverse transcription step of HIV (and also hepadnaviruses [Renard et al., 2010]), *APOBEC3G* induces multiple G → A mutations in the viral genome, many of which will be deleterious and so inhibit viral function. However, HIV-1 has evolved an anti-*APOBEC3G* response controlled by the *vif* gene (Sheehy et al., 2002). As the mutational signatures of *APOBEC3G* action have also been observed in human endogenous retroviruses, it is possible that *APOBEC3G* has been functioning as an antiretroviral agent for millions of years (Armitage et al., 2008).

The intensive and intricate nature of the arms race between host and virus may in part explain why estimates of deep divergence times using currently circulating exogenous viruses perform so badly. Specifically, if there is a strong arms race between host and virus, and if, because of the intrinsic constraints that act on virus proteins, only a limited number of amino acid changes in the virus are permitted, then this greatly restricted number of evolutionary pathways may lead to a high number of multiple substitutions at single sites, in turn resulting in the inaccurate measurement of evolutionary distances. Indeed, convergent evolution, one manifestation of the limited number of evolutionary pathways open to viruses, appears to be especially common in RNA viruses (Bull et al., 1997; Cuevas et al., 2002; Holmes, 2003b).

The discovery and characterization of endogenous viral elements has opened up an important new avenue of research into virus evolution, raising important questions on how such elements are generated, whether they can sometimes contribute beneficial functions to the host cell, and the long-term evolutionary history of their exogenous relatives. Perhaps paradoxically, given that they reside in host genomes, the most profound impact of endogenous viruses may be on our understanding of the time scale of viral evolution, and highlighting the need for phylogenetic methods that are better suited to the analysis of highly divergent sequences. Endogenous viruses have therefore told us that there is still a great deal to learn about their exogenous relatives.

REFERENCES

Armitage, A.E., Katzourakis, A., de Oliveira, T., Welch, J.J., Belshaw, R., Bishop, K.N., Kramer, B., McMichael, A.J., Rambaut, A., and Iversen, A.K. (2008). Conserved footprints of APOBEC3G on hypermutated human immunodeficiency virus type 1 and human endogenous retrovirus HERV-K(HML2) sequences. J. Virol. *82*, 8743–8761.

Arnaud, F., Caporale, M., Varela, M., Biek, R., Chessa, B., Alberti, A., Golder, M., Mura, M., Zhang, Y.P., Yu, L., et al. (2007). A paradigm for virus-host coevolution: sequential counter-adaptations between endogenous and exogenous retroviruses. PLoS Pathog. *3*, e170.

Barrette, R.W., Xu, L., Rowland, J.M., and McIntosh, M.T. (2011). Current perspectives on the phylogeny of *Filoviridae*. Infect. Genet. Evol., in press. Published online June 30, 2011.

Bejarano, E.R., Khashoggi, A., Witty, M., and Lichtenstein, C. (1996). Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. Proc. Natl. Acad. Sci. USA *93*, 759–764.

Belyi, V.A., Levine, A.J., and Skalka, A.M. (2010a). Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the *parvoviridae* and *circoviridae* are more than 40 to 50 million years old. J. Virol. *84*, 12458–12462.

Belyi, V.A., Levine, A.J., and Skalka, A.M. (2010b). Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. PLoS Pathog. *6*, e1001030.

Benveniste, R.E., and Todaro, G.J. (1974). Evolution of C-type viral genes: inheritance of exogenously acquired viral genes. Nature *252*, 456–459.

Bernard, H.-U., Chan, S.-Y., and Delius, H. (1994). Evolution of papillomaviruses. Curr. Top. Microbiol. Immunol. *186*, 33–54.

Bézier, A., Annaheim, M., Herbinière, J., Wetterwald, C., Gyapay, G., Bernard-Samain, S., Wincker, P., Roditi, I., Heller, M., Belghazi, M., et al. (2009). Polydnaviruses of braconid wasps derive from an ancestral nudivirus. Science *323*, 926–930.

Bonilla Guerrero, R., and Roberts, L.R. (2005). The role of hepatitis B virus integrations in the pathogenesis of human hepatocellular carcinoma. J. Hepatol. *42*, 760–777.

Bull, J.J., Badgett, M.R., Wichman, H.A., Huelsenbeck, J.P., Hillis, D.M., Gulati, A., Ho, C., and Molineux, I.J. (1997). Exceptional convergent evolution in a virus. Genetics *147*, 1497–1507.

Crochu, S., Cook, S., Attoui, H., Charrel, R.N., De Chesse, R., Belhouchet, M., Lemasson, J.J., de Micco, P., and de Lamballerie, X. (2004). Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes. J. Gen. Virol. *85*, 1971–1980.

Cuevas, J.M., Elena, S.F., and Moya, A. (2002). Molecular basis of adaptive convergence in experimental populations of RNA viruses. Genetics *162*, 533–542.

Denison, M.R., Graham, R.L., Donaldson, E.F., Eckerle, L.D., and Baric, R.S. (2011). Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. RNA Biol. 8, 270–279.

Drummond, A.J., Pybus, O.G., Rambaut, A., Forsberg, R., and Rodrigo, A.G. (2003). Measurably evolving populations. Trends Ecol. Evol. 18, 481–488.

Duffy, S., Shackelton, L.A., and Holmes, E.C. (2008). Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. 9, 267–276.

Elde, N.C., Child, S.J., Geballe, A.P., and Malik, H.S. (2009). Protein kinase R reveals an evolutionary model for defeating viral mimicry. Nature 457, 485–489.

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

Evans, D.T., Serra-Moreno, R., Singh, R.K., and Guatelli, J.C. (2010). BST-2/ tetherin: a new component of the innate immune response to enveloped viruses. Trends Microbiol. 18, 388–396.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. Nat. Rev. Genet. 9, 397–405.

Frank, A.C., and Wolfe, K.H. (2009). Evolutionary capture of viral and plasmid DNA by yeast nuclear chromosomes. Eukaryot. Cell 8, 1521–1531.

Geuking, M.B., Weber, J., Dewannieux, M., Gorelik, E., Heidmann, T., Hengartner, H., Zinkernagel, R.M., and Hangartner, L. (2009). Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. Science 323, 393–396.

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. USA 105, 20362–20367.

Gilbert, C., and Feschotte, C. (2010). Genomic fossils calibrate the long-term evolution of hepadnaviruses. PLoS Biol. 8, e1000495.

Hahn, B.H., Shaw, G.M., De Cock, K.M., and Sharp, P.M. (2000). AIDS as a zoonosis: scientific and public health implications. Science 287, 607–614.

Hamilton, W.D., Axelrod, R., and Tanese, R. (1990). Sexual reproduction as an adaptation to resist parasites (a review). Proc. Natl. Acad. Sci. USA 87, 3566–3573.

Holmes, E.C. (2003a). Molecular clocks and the puzzle of RNA virus origins. J. Virol. 77, 3893–3897.

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

Holmes, E.C. (2009). The Evolution and Emergence of RNA Viruses. Oxford Series in Ecology and Evolution (OSEE) (Oxford: Oxford University Press).

Holmes, E.C. (2011). What does virus evolution tell us about virus origins? J. Virol. 85, 5247–5251.

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

Johnson, W.E. (2010). Endless forms most viral. PLoS Genet. 6, e1001210.

Kapoor, A., Simmonds, P., and Lipkin, W.I. (2010). Discovery and characterization of mammalian endogenous parvoviruses. J. Virol. 84, 12628–12635.

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

Katzourakis, A., and Tristem, M. (2005). Phylogeny of human endogenous and exogenous retroviruses. In Retroviruses and Primate Genome Evolution, E.D. Sverdlov, ed. (Austin, TX: Landes Bioscience), pp. 186–203.

Katzourakis, A., Tristem, M., Pybus, O.G., and Gifford, R.J. (2007). Discovery and analysis of the first endogenous lentivirus. Proc. Natl. Acad. Sci. USA 104, 6261–6265.

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

Keckesova, Z., Ylinen, L.M., Towers, G.J., Gifford, R.J., and Katzourakis, A. (2009). Identification of a RELIK orthologue in the European hare (Lepus europaeus) reveals a minimum age of 12 million years for the lagomorph lentiviruses. Virology 384, 7–11.

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, e1000462.

Kitchen, A., Shackelton, L.A., and Holmes, E.C. (2011). Family level phylogenies reveal modes of macroevolution in RNA viruses. Proc. Natl. Acad. Sci. USA 108, 238–243.

Koonin, E.V., Senkevich, T.G., and Dolja, V.V. (2006). The ancient Virus World and evolution of cells. Biol. Direct 1, 29.

Leroy, E.M., Kumulungui, B., Pourrut, X., Rouquet, P., Hassanin, A., Yaba, P., Délicat, A., Paweska, J.T., Gonzalez, J.P., and Swanepoel, R. (2005). Fruit bats as reservoirs of Ebola virus. Nature 438, 575–576.

Liu, W., Worobey, M., Li, Y., Keele, B.F., Bibollet-Ruche, F., Guo, Y., Goepfert, P.A., Santiago, M.L., Ndjango, J.B., Neel, C., et al. (2008). Molecular ecology and natural history of simian foamy virus infection in wild-living chimpanzees. PLoS Pathog. 4, e1000097.

Lynch, M., and Conery, J.S. (2003). The origins of genome complexity. Science 302, 1401–1404.

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.

Medstrand, P., van de Lagemaat, L.N., and Mager, D.L. (2002). Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome Res. 12, 1483–1495.

Meyerson, N.R., and Sawyer, S.L. (2011). Two-stepping through time: mammals and viruses. Trends Microbiol. 19, 286–294.

Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S., et al. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature 403, 785–789.

Minskaia, E., Hertzig, T., Gorbalenya, A.E., Campanacci, V., Cambillau, C., Canard, B., and Ziebuhr, J. (2006). Discovery of an RNA virus 3′->5′ exoribonuclease that is critically involved in coronavirus RNA synthesis. Proc. Natl. Acad. Sci. USA 103, 5108–5113.

Murcia, P.R., Arnaud, F., and Palmarini, M. (2007). The transdominant endogenous retrovirus enJS56A1 associates with and blocks intracellular trafficking of Jaagsiekte sheep retrovirus Gag. J. Virol. 81, 1762–1772.

Nisole, S., Stoye, J.P., and Saïb, A. (2005). TRIM family proteins: retroviral restriction and antiviral defence. Nat. Rev. Microbiol. 3, 799–808.

Nozawa, M., Suzuki, Y., and Nei, M. (2009). Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. Proc. Natl. Acad. Sci. USA 106, 6700–6705.

Ohta, T. (1992). The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. 23, 263–286.

Parrish, C.R. (1990). Emergence, natural history, and variation of canine, mink, and feline parvoviruses. Adv. Virus Res. 38, 403–450.

Plyusnin, A., and Morzunov, S.P. (2001). Virus evolution and genetic diversity of hantaviruses and their rodent hosts. Curr. Top. Microbiol. Immunol. 256, 47–75.

Ramsden, C., Melo, F.L., Figueiredo, L.M., Holmes, E.C., and Zanotto, P.M.de A.; VGDN Consortium. (2008). High rates of molecular evolution in hantaviruses. Mol. Biol. Evol. 25, 1488–1492.

Ramsden, C., Holmes, E.C., and Charleston, M.A. (2009). Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. Mol. Biol. Evol. 26, 143–153.

Renard, M., Henry, M., Guétard, D., Vartanian, J.P., and Wain-Hobson, S. (2010). APOBEC1 and APOBEC3 cytidine deaminases as restriction factors for hepadnaviral genomes in non-humans in vivo. J. Mol. Biol. 400, 323–334.

Roiz, D., Vázquez, A., Seco, M.P., Tenorio, A., and Rizzoli, A. (2009). Detection of novel insect flavivirus sequences integrated in *Aedes albopictus* (Diptera: Culicidae) in Northern Italy. Virol. J. *6*, 93.

Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brottier, P., Camus, J.C., Cattolico, L., et al. (2002). Genome sequence of the plant pathogen *Ralstonia solanacearum*. Nature *415*, 497–502.

Sanjuán, R., Nebot, M.R., Chirico, N., Mansky, L.M., and Belshaw, R. (2010). Viral mutation rates. J. Virol. *84*, 9733–9748.

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.

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. USA *102*, 379–384.

Sharp, P.M., Bailes, E., Chaudhuri, R.R., Rodenburg, C.M., Santiago, M.O., and Hahn, B.H. (2001). The origins of acquired immune deficiency syndrome viruses: where and when? Philos. Trans. R. Soc. Lond. B Biol. Sci. *356*, 867–876.

Sheehy, A.M., Gaddis, N.C., Choi, J.D., and Malim, M.H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature *418*, 646–650.

Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J., and Gorbalenya, A.E. (2003). Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J. Mol. Biol. *331*, 991–1004.

Stremlau, M., Owens, C.M., Perron, M.J., Kiessling, M., Autissier, P., and Sodroski, J. (2004). The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature *427*, 848–853.

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.

Tanne, E., and Sela, I. (2005). Occurrence of a DNA sequence of a non-retro RNA virus in a host plant genome and its expression: evidence for recombination between viral and host RNAs. Virology *332*, 614–622.

Tarlinton, R.E., Meers, J., and Young, P.R. (2006). Retroviral invasion of the koala genome. Nature *442*, 79–81.

Taylor, D.J., and Bruenn, J. (2009). The evolution of novel fungal genes from non-retroviral RNA viruses. BMC Biol. *7*, 88.

Taylor, D.J., Leach, R.W., and Bruenn, J. (2010). Filoviruses are ancient and integrated into mammalian genomes. BMC Evol. Biol. *10*, 193.

The MHC Sequencing Consortium. (1999). Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. Nature *401*, 921–923.

van der Loo, W., Abrantes, J., and Esteves, P.J. (2009). Sharing of endogenous lentiviral gene fragments among leporid lineages separated for more than 12 million years. J. Virol. *83*, 2386–2388.

Weiss, R.A., Mason, W.S., and Vogt, P.K. (1973). Genetic recombinants and heterozygotes derived from endogenous and exogenous avian RNA tumor viruses. Virology *52*, 535–552.

Wertheim, J.O., and Kosakovsky Pond, S.L. (2011). Purifying selection can obscure the ancient age of viral lineages. Mol. Biol. Evol., in press. Published online June 24, 2011.

Wertheim, J.O., and Worobey, M. (2009). Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. PLoS Comput. Biol. *5*, e1000377.

Worobey, M., Telfer, P., Souquière, S., Hunter, M., Coleman, C.A., Metzger, M.J., Reed, P., Makuwa, M., Hearn, G., Honarvar, S., et al. (2010). Island biogeography reveals the deep history of SIV. Science *329*, 1487.

Yang, Z., and Bielawski, J.P. (2000). Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. (Amst.) *15*, 496–503.

Zhou, Y., and Holmes, E.C. (2007). Bayesian estimates of the evolutionary rate and age of hepatitis B virus. J. Mol. Evol. *65*, 197–205.