Studies of endogenous retroviruses reveal a continuing evolutionary saga

Jonathan P. Stoye

Abstract | Retroviral replication involves the formation of a DNA provirus integrated into the host genome. Through this process, retroviruses can colonize the germ line to form endogenous retroviruses (ERVs). ERV inheritance can have multiple adverse consequences for the host, some resembling those resulting from exogenous retrovirus infection but others arising by mechanisms unique to ERVs. Inherited retroviruses can also confer benefits on the host. To meet the different threats posed by endogenous and exogenous retroviruses, various host defences have arisen during evolution, acting at various stages on the retrovirus life cycle. In this Review, I describe our current understanding of the distribution and architecture of ERVs, the consequences of their acquisition for the host and the emerging details of the intimate evolutionary relationship between virus and vertebrate host.

Tumorigenic
Capable of forming tumours. Some but not all retroviruses are capable of changing cell growth properties, resulting in cancer. The kinds of tumours seen include carcinomas, sarcomas and leukaemias.

Provirus
The DNA form of a retrovirus integrated into the genomes of retrovirus-infected cells or organisms. Coding sequences are flanked by long terminal repeats.

siRNA screens
(Small interfering RNA screens). Widely used 'knockout' studies of gene function that use siRNAs, which are double-stranded RNA molecules of 20–25 nucleotides in length that are capable of interfering with the expression of RNA.

Division of Virology, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.
e-mail: jstoye@nimr.mrc.ac.uk
doi:10.1038/nrmicro2783
Published online 8 May 2012

With their unique life cycle, tumorigenic properties and, more recently, their role in the AIDS epidemic, retroviruses have attracted enormous attention over the past 50 years[1,2]. However, studies of these properties have perhaps overshadowed a number of observations pointing to an equally interesting facet of retroviral biology: the intimate relationship and constant evolutionary interplay between retroviruses and their vertebrate hosts, which probably dates back hundreds of millions of years and persists to the present day.

The replication cycle of retroviruses (FIG. 1) is best known for the involvement of two unique, virally encoded enzymes, reverse transcriptase and integrase[3]. They permit the conversion of RNA into DNA, followed by the integration of viral DNA into the host genome, forming a DNA provirus. In addition, studies of specific steps in the viral life cycle[3], as well as large-scale siRNA screens, have shown that multiple cellular factors are absolutely required for virus replication[4].

As retroviruses do not usually lyse their target cell, integration allows a long-term association between cell and virus. If the infected cell is a germ cell, colonization of the germ line by the virus is possible. Numerous studies indicate that such events have occurred multiple times during the course of evolution. Such inherited proviruses are called endogenous retroviruses (ERVs), and they provide a 'fossil record' of past retroviral infections dating back many millions of years, which suggests a never-ending stream of retroviral challenges for vertebrates[5].

In response to such infections, several host defences giving rise to intrinsic immunity to retroviral infection have evolved. In turn, further changes to retroviruses have arisen. This Review traces the long-standing relationship between retroviruses and their hosts, which was initially revealed by the study of ERVs, and discusses some of the resulting changes driven by their evolutionary conflict.

ERVs

ERVs were originally discovered in the late 1960s as inherited cellular genes that could encode retroviral gene products or replication-competent retroviruses[6]. Combining virological and immunological techniques with Mendelian genetics led, almost simultaneously, to the discovery of endogenous avian leukosis virus (ALV) in birds as well as murine leukaemia virus (MLV) and mouse mammary tumour virus (MMTV) in mice. The presence of inherited viral genomes was then confirmed in hybridization experiments. The involvement of these or related proviruses in animal models of cancer[7] prompted the search for more ERVs, particularly in humans. Numerous ERVs were identified using low-stringency hybridization[8] or PCR strategies designed to amplify conserved viral sequences[9]; others were found serendipitously following the unexpected recovery of replicating virus[10]. The extent of ERV presence has only been fully appreciated with the completion of full genome sequences from a range of organisms, including humans, mice and chickens[11–13], and the development of increasingly sophisticated *in silico* methods for ERV identification[14].

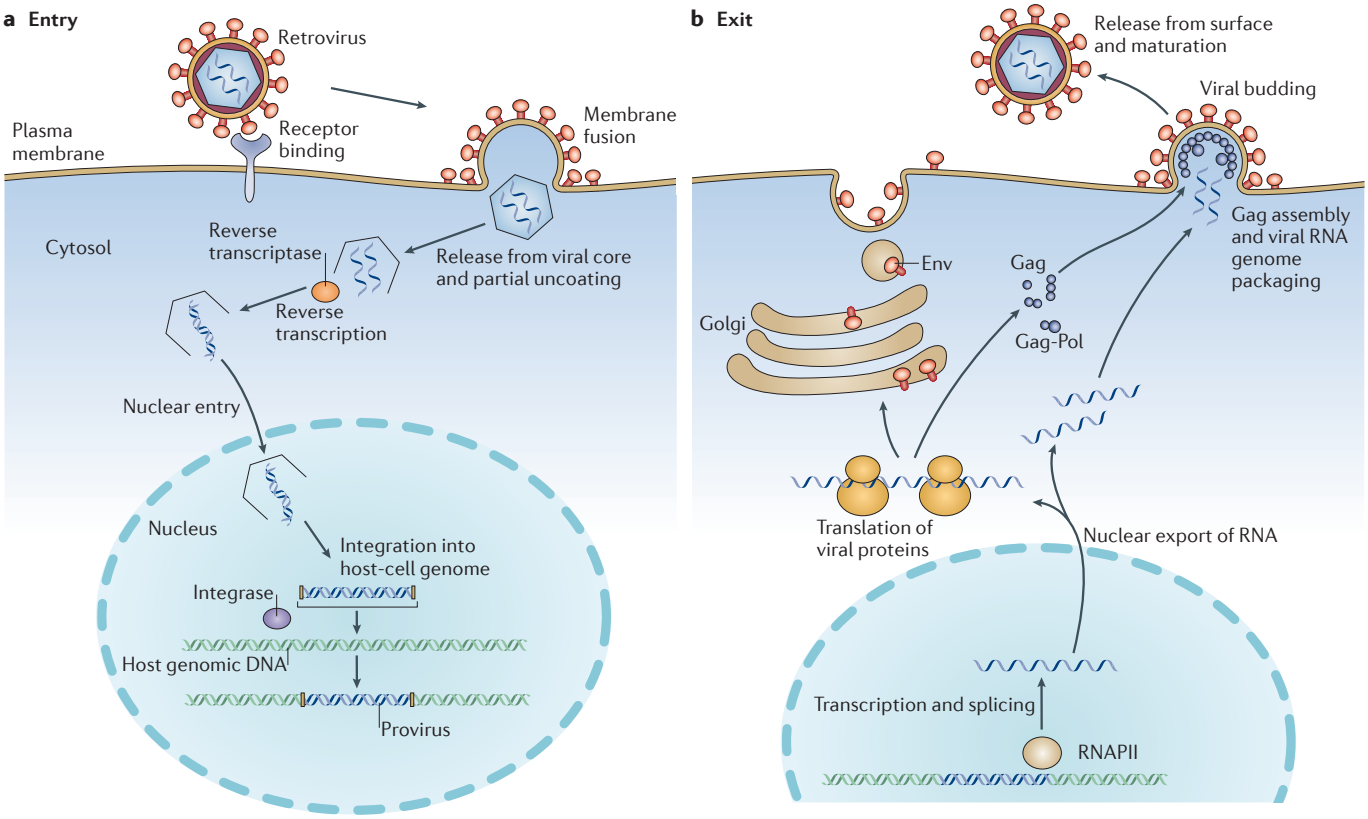

Figure 1 | **Steps in the retroviral life cycle.** Different events in the life cycle of retroviruses are illustrated. **a** | Viral entry into cells involves the following steps: binding to a specific receptor on the cell surface; membrane fusion either at the plasma membrane or from endosomes (not shown); release of the viral core and partial uncoating; reverse transcription; transit through the cytoplasm and nuclear entry; and integration into cellular DNA to give a provirus. **b** | Viral exit involves the following steps: transcription by RNA polymerase II (RNAPII); splicing and nuclear export of viral RNA; translation of viral proteins, Gag assembly and RNA packaging; budding through the cell membrane; and release from the cell surface and virus maturation. A more complete description of these events can be found in REF. 3. Figure is modified, with permission, from REF. 155 © (2007) Macmillan Publishers Ltd. All rights reserved.

Somatic cells
Differentiated cells of the body that lack potential to contribute to the germ line.

Solo LTRs
(Solo long terminal repeats). Lone LTRs in the genome. Homologous recombination between the two LTRs of a provirus results in excision of most of the provirus, leaving behind a solitary LTR in the genome at the site of the previous provirus.

Retrotransposons
Genetic elements that can increase in copy numbers by a mechanism involving reverse transcription of an RNA intermediate followed by integration into the genome.

Long interspersed nuclear elements
(LINEs). Long retrotransposons that encode reverse transcriptases but show genetic organizations and modes of amplification that are different from those of retroviruses; in particular, they lack long terminal repeats.

Phylogenetic analysis revealed that the human genome contains at least 31 distinct groups of ERVs, ranging in copy number from one to many thousands[15,16]. Similar distributions are found in other species[17,18]. These numbers mean that ERV classification and nomenclature present substantial difficulties[19] (BOX 1).

ERV genetic architecture. Individual ERVs are present in a form that is indistinguishable from the proviruses that result from retroviral infection of somatic cells (FIG. 2). They comprise two long terminal repeats (LTRs), 300–1,200 nucleotides in length, separated by approximately 5–10 kb of sequence encoding the canonical retroviral *gag*, *pol* and *env* genes. Retroviral integration is essentially random[3]; thus, each ERV is present at a unique chromosomal location, and proviruses with very similar sequences can be distinguished by virtue of their host flanking sequences[5]. Most ERVs are clearly defective, carrying a multitude of inactivating mutations in their coding sequences that vary in size from single nucleotide changes to large deletions. Some mutations seem to arise by apolipoprotein B mRNA-editing complex 3 (APOBEC3)-mediated editing before integration[20], but most accumulate after endogenization. Accordingly, infectious ERVs have not been identified in human DNA, suggesting that the accumulated mutations have rendered human ERVs inactive. By contrast, in other organisms, including mice and cats, more intact ERVs have been found and infectious ERVs identified[17]. In all species, the number of full-length proviruses is at least tenfold lower than the number of 'solo LTRs' (REF. 21) that are thought to result from recombination between the two LTRs of an integrated provirus[22]. ERVs encoding accessory genes characteristic of complex exogenous retroviruses like HIV-1 are rare and have only been reported in a small number of species[23–25].

Analyses of completed genome sequences have revealed that between four and ten per cent of vertebrate DNA is made up of retroviral remnants[18,26]. This amount is several times higher than that corresponding to sequences encoding genes[11] but is substantially less than that derived from non-LTR-containing retrotransposons (other transposable elements with an RNA intermediate but with a genetic organization different from that of retroviruses), such as long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs)[27].

Box 1 | **ERV classification and nomenclature**

The nomenclature system for endogenous retroviruses (ERVs) has developed in a manner reflecting the history of their discovery. It first became common to name endogenous proviruses after the most closely related exogenous retrovirus, such as murine leukaemia virus (MLV), as well as subgroups, such as xenotropic MLV (XMV). It is now standard to add one or two letters before the designation ERV to indicate the species in which they were initially identified; thus, HERV indicates an ERV first seen in human DNA and MERV or MuERV implies one originally found in mice. HERVs are further classified on the basis of the tRNA that binds to the viral primer-binding site (PBS) to prime reverse transcription. Hence, HERV-K implies a provirus or groups of proviruses that use a lysine tRNA, no matter their relationship to one another. In some cases, the PBS sequence was not available when novel elements were first discovered, leading to the names based on neighbouring genes (for example, HERV.ADP), clone number (for example, HERV.S71) or amino acid motifs (for example, HERV.FRD). Additional designations based on the probe used for cloning (for example, HERV.HML) and sub-divisions based on the degree of sequence identity in reverse transcriptase or phylogenetic reconstructions (for example, HERV-K(HML-2)) have also been used. More recently, phylogenetic methods have been used to define different groups of sequences.

At another level, it is also often necessary to distinguish among specific proviruses at discrete chromosomal locations in a group, and several different systems have evolved for this purpose. Most commonly, individual proviruses are simply numbered, as with XMV1 and HERV-K 108. In the case of HERVs, some investigators have chosen to use cytogenetic designations to distinguish among related proviruses, as in HERV-K 11q22. As noted above, neighbouring genes have also provided a basis for distinction between specific proviruses.

LINEs and SINEs lack an extracellular phase in their life cycles and are thus not associated with transmissible disease. Consequently, they have received somewhat less attention than ERVs. Together, these products of reverse transcription make up nearly half of the vertebrate genome[28] and therefore represent a significant challenge to developing whole-genome sequence assemblies using current short-read sequencing techniques. Given their undoubted biological importance, it would be regrettable if such sequences were therefore ignored in resequencing projects.

It should also be noted that a small number of integration events involving RNA viruses other than retroviruses have been reported. These include bornavirus and filovirus[29]. Such viruses do not normally undergo reverse transcription; it seems likely that such integrations took place following reverse transcription mediated by a LINE element or after illegitimate recombination with an ERV. Although some of these integrations contain open reading frames, their evolutionary significance for their hosts, if any, is unclear. Conversely, their analysis has clear implications for our understanding of the timing of virus evolution (see REF. 29 for a review).

Germline colonization and ERV distribution. Germline colonization has been observed both in nature and in the laboratory. A spreading endogenization of a virus similar to gibbon ape leukaemia virus but probably of murine origin is currently taking place in Australian koalas[30]. Infection of early stage mouse embryos with Moloney MLV resulted in the formation of newly integrated proviruses that could then be transmitted through the germ line[31]. Similar results were obtained by injecting recombinant ALV into fertilized eggs[32]. Further studies identified inbred strains of mice that are spontaneously acquiring novel proviruses by a process involving infection of oocytes in the ovaries of viraemic females[33]. The initial viraemia can reflect replication of either an exogenous retrovirus or an ERV, thereby mimicking germline colonization and ERV amplification, respectively.

We can thus infer that the current distribution of ERVs in different species reflects a number of independent germline colonization events followed by proviral amplification resulting from further infection or other retrotranspositional events[5,34]. The increase in provirus number will cease when the individual founder proviruses and their infectious progeny accumulate sufficient random mutations to make them non-infectious, a process common to all non-selected genomic sequences. Alternatively, amplification will cease if the host develops a means of restricting virus replication or if homologous recombination between LTRs removes the body of the founder provirus.

In conclusion, these analyses have shown that retroviruses, with essentially the same genetic organization as modern day viruses, have coexisted with our ancestors for a very lengthy period (BOX 2). Unfortunately, however, they do not tell us what fraction of the retroviruses that have ever existed have been sampled, where they came from or the effectiveness of host defences against cross-species infection and the formation of ERVs.

Consequences of ERV acquisition

Retroviral colonization of the germ line can have a range of consequences for the host organism. As might be expected, many are detrimental. However, several are not, and at least one (see discussion of syncytin below) seems to have had a profound effect on vertebrate evolution. Mechanistically, we can distinguish three groups of effects: those caused by the expression of viruses or viral proteins; those resulting from the action of viral sequences controlling RNA transcription, splicing or stability; and those caused by the insertion of bulk DNA.

Expression of viral proteins. Exogenous retroviruses are clearly responsible for a range of pathological conditions, such as cancer and AIDS. This has prompted a multitude of studies to investigate the pathogenic potential of ERVs. However, with a few notable exceptions, such as the mammary and lymphoid tumours of certain inbred mice[7], replication-competent viruses of endogenous origin have not been isolated from diseased tissues, and the potential roles of ERVs in conditions such as cancer, autoimmunity and various neurological conditions remain, at best, unproven hypotheses[35–38]. Distinguishing cause and effect can be a substantial problem because, in many cases, ERV transcription may simply reflect changes in the amount or nature of cellular transcription factors present under altered physiological conditions. Further, immune responses to ERVs may reflect their availability as targets rather than any essential role in disease induction. Nevertheless, it remains possible that overexpression of the Rec protein of human ERV-K (HERV-K) may have a role in the formation of germ cell tumours in humans by derepressing oncogenic

Short interspersed nuclear elements
(SINEs). Short retrotransposons with no coding sequences. They can be reverse transcribed by LINE-encoded reverse transcriptases.

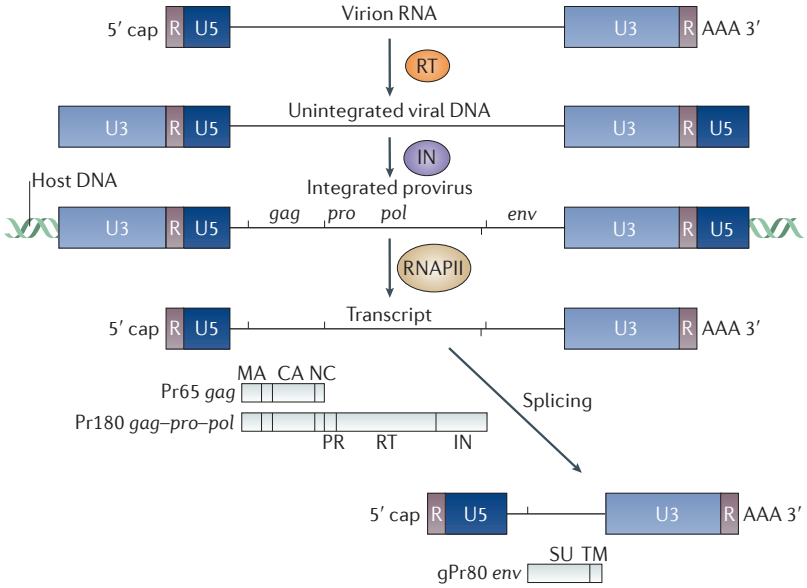

Figure 2 | **Retroviral structures.** Different forms of retroviral genetic information. Retroviruses contain RNA during their extracellular phase. After cell infection, viral RNA is reverse transcribed into double-stranded DNA and the long terminal repeats (LTRs) are created. Viral DNA is subsequently integrated into the host cell's genomic DNA to form a provirus that is later transcribed by host cell RNA polymerase II (RNAPII). The typical simple retrovirus has three genes: *gag*, *pol* and *env*. The *gag* gene encodes the structural proteins that make up the viral core, matrix (MA), capsid (CA) and nucleocapsid (NC), and these are initially synthesized as a single polyprotein. The *pol* gene encodes the viral enzymes protease (PR), reverse transcriptase (RT) and integrase (IN). Initial translation is as a Gag–Pol fusion protein from genome-length RNA. The *env* gene encodes the surface (SU) and transmembrane (TM) proteins responsible for receptor binding and membrane fusion. The Env precursor protein is translated from spliced subgenomic DNA. The relative sizes of the precursor proteins shown in this figure correspond to Moloney murine leukaemia virus. Complex retroviruses can encode up to six accessory proteins. They are generally located at (and sometimes overlapping) the *pol*–*env* and *env*–LTR boundaries.

transcription factors[39]. In some cases (for example, feline leukaemias[40]), individual ERV genes may have an important role in disease induction by a process involving recombination with an exogenous virus and activation of cellular oncogenes[7]. In retrospect, the relative absence of directly pathogenic viruses from the germ line is perhaps not that surprising, as they would represent a substantial negative evolutionary pressure. It is also worth noting that one prime example of ERV-associated tumorigenesis, the thymomas that arise spontaneously in nearly all AKR mice, is a highly artificial case. AKR mice were originally derived in the 1930s following multiple generations of inbreeding accompanied by selection for high tumour incidence. Detailed virological analysis revealed that the disease process involves recombination between at least three different types of endogenous MLV[41] to generate the agent responsible for disease. The same pattern of recombination is repeated in each animal.

Control of host processes by ERV sequences. The retroviral provirus carries specific sequences that control the synthesis and processing of its own RNA. These sequences act as signals to cellular proteins that carry out the transcription, splicing and polyadenylation of viral RNA (FIG. 3a). They can also affect the expression of genes neighbouring the viral integration site. Each kind of signal can participate in tumorigenesis by exogenous viruses (FIG. 3b), and all might be expected to act in an endogenous context to activate transcription or interfere with proper processing of a cellular mRNA. In practice, mutagenesis following insertion within an intron of a cellular gene is quite common, at least in mice[42], when utilization of a viral splice acceptor site can lead to mis-splicing and the formation of a cell–virus hybrid mRNA. Effects associated with promoter or enhancer insertion are less common[43], although not without precedent. For example, some data suggest that aberrant expression of the colony-stimulating factor 1 receptor (*CSF1R*) proto-oncogene in some human lymphomas is driven by activation of an endogenous LTR[44]. However, integrations leading to uncontrolled overexpression of most cellular transcripts are likely to perturb normal development and will be negatively selected. Mutations resulting in mis-splicing have been observed to revert following loss of the provirus by homologous recombination across the viral LTRs[22]; nevertheless, uncontrolled ERV amplification might be expected to have serious mutational consequences[45].

Given that recombination occurs between the LTRs of individual proviruses, one might predict that a similar process might occur between different, related proviruses, leading to chromosome instability during evolution[46]. In practice, such events seem to happen only rarely[47], although reports of male infertility associated with recombination-mediated Y-chromosome deletions indicate that such processes can occur[48]. Deep sequencing of tumours may yet reveal the potential importance of such events in the course of tumour development.

Cost–benefit balance. One might imagine that a further potential problem associated with ERV inheritance would result from the need to replicate so much 'unnecessary' DNA or to express unneeded proteins, which presumably impart a substantial metabolic burden on the organism, although any fitness costs are hard to quantify. As such, a gradual increase in ERV copy number during evolution may have been accompanied by a gradual increase in the metabolic capacity of their hosts to meet this need.

One reason for the host to accept any burdens associated with the presence of ERVs would be if ERVs could be exploited to 'pay their way': that is, if their presence is beneficial. For instance, ERVs might contribute resistance to exogenous retroviruses (variations on this theme are explored below). Alternatively, retroviral gene products might be co-opted to play a part in normal development. One striking example is provided by the case of the syncytin genes, for which compelling evidence has been presented suggesting that ERV Env proteins have a vital role in placenta formation[49,50]. Such a role might arise by virtue of the cell-fusion properties of ERV Env proteins, leading to the production of the syncytiotrophoblast layer, or because they have immunosuppressive properties that help to prevent fetal rejection by the mother. It is possible that both properties are important[51]. Such

Polyadenylation
mRNAs are characterized by a stretch of adenine residues at their 3′ termini. Addition of these poly(A) tails, polyadenylation, is an important step in mRNA maturation and is signalled by specific nucleotide sequences.

Splice acceptor
A sequence element that is important for RNA splicing. RNA splicing is a vital step in RNA maturation in which coding exons are joined by intron removal. It is signalled by specific splice donor and splice acceptor sequences.

Box 2 | **Dating integration events**

Two distinct properties of retroviral replication allow the dating of individual proviruses. First, the mechanism of reverse transcription dictates that the sequences of the two long terminal repeats (LTRs) of a given provirus are identical at the time of integration. Determining the current differences in sequence allows the relative dating of members of particular groups of endogenous retrovirus (ERV); applying a rate for neutral sequence variation permits a calculation of the time since integration occurred[148]. Estimates can be distorted either by differential mutation rates dependent on the integration site[149] or by gene conversion events between the LTRs of different group members[46]. Second, as retroviral integration is essentially random, a provirus with identical host flanking sequences in two different species can be taken as the result of one integration event that predated the separation of the two species[146]. Indeed, such proviruses can be used as tags for developing improved evolutionary trees. It is noteworthy that such proviruses are more closely related than different members of the same group of sequences in a one species, an observation that complicates the development of a rational nomenclature system for ERVs. Use of either dating method allows the conclusion that retroviruses have been integrating into the ancestors of all extant vertebrate species for tens, if not hundreds, of millions of years, although random mutational change complicates the identification of very ancient vertebrate retroviruses. Retroviral endogenization seems to be an ongoing process: whereas some viruses first appeared many years ago, others seem to have arisen much more recently[150]. Thus, most groups of human ERVs are present in Old World monkeys and apes, suggesting that these colonization events occurred at least 30 million years ago[15,151,152]. By contrast, two groups of chimpanzee virus are not present in humans, implying that they entered the chimpanzee genome in the last five million years[26]. In some viral groups, such as human ERV-K (HERV-K), viral amplification has continued for lengthy periods[153], raising questions about their survival, in active form, in the face of random mutation and the absence of purifying selection. Complementation between two defective proviruses followed by recombination and/or bursts of replication as exogenous viruses might provide mechanisms for this persistence[150,154].

assimilation has occurred independently, and involving ERVs of different groups, in at least three different orders of mammals: Primates, Rodentia and Lagomorpha[52].

ERV transcriptional control elements, such as promoters or enhancers, may also be used to control host gene transcription. For example, the insertion of a provirus upstream of a duplicated copy of the pancreatic amylase gene in humans led to the expression of amylase in saliva and the consequent ability to digest starch in the mouth[53], which had profound consequences for agricultural development. In another example, a provirus present in only humans and great apes controls the testis-specific expression of germ cell-associated transcriptionally active p63 (GTAp63), a protein that protects male germ cells from cancer[54].

Clearly, the presence of ERVs has associated costs. Integrations with severe costs, such as ablation of essential genes or rapid induction of disease, will not be fixed in the population. However, ERVs can bring benefits, and provided that inherited ERV replication can be controlled, the association can be stably maintained. This has led to the evolution of different mechanisms regulating retrovirus activity.

Controlling retrovirus replication

Recently, there has been increasing interest in understanding the many host factors that can influence the replication of endogenous and exogenous retroviruses. They act at multiple stages of the viral life cycle, in both intracellular and extracellular phases (FIG. 4). They can be divided into three groups. First are a group of cellular proteins that can act to silence or repress viral RNA transcription by epigenetic mechanisms[55–57]. Second are the dedicated restriction factors, proteins whose primary purpose seems to be to act as a first line of defence against viral infection[58]. If expressed in somatic cell hybrids, these proteins will act as dominant inhibitors of retrovirus replication and spread. Third are proteins that perform 'normal' functions in the cell but are also exploited by retroviruses during the course of viral replication[4,59]: for example, the receptor necessary for initial virus binding to the cell[60]. Subject to the constraints of their normal cellular function, such proteins can mutate to prevent interaction with the virus or change the intracellular environment, leading to a block in virus replication. Genetically, such factors will act in recessive fashion. Some examples of the host factors that affect retroviral replication are given below.

Epigenetic mechanisms. Heritable changes in gene expression without changes in underlying DNA sequence can occur by alteration of DNA methylation or histone modification, thereby providing a means for global control of ERV expression. It has been clear for several years that ERV transcription can be controlled by the methylation state of genomic DNA[56]. Treatment of cells with agents that promote the demethylation of genomic DNA[61], such as 5-azacytidine, results in ERV induction. At the same time, study of newly integrated Moloney MLV proviruses in embryonal carcinoma cells and of newly formed endogenous Moloney MLVs resulting from embryo infection revealed a correlation between virus expression and methylation state[62]. These findings, together with the heritable nature of methylation patterns[63], suggested that methylation of genomic DNA is an excellent means of controlling genomic parasites such as the ERVs and led to the hypothesis that this may have been the principle driving force behind the evolution of cytosine methylation as a means for gene regulation[64]. However, it remains unclear how novel ERVs are initially targeted for silencing, although it seems likely that this process may involve a protein complex comprising tripartite motif-containing 28 (TRIM28) and one or more zinc-finger proteins[65,66]. TRIM28 is one member of a family of proteins characterized by the presence of an amino-terminal RBCC (ring, B box and coiled-coil) domain, many of which may be associated with resistance to viral infection[67].

Although methylation provides a mechanism for ERV silencing in adult tissues, it cannot by itself explain the control of ERV sequences during early development, when zygotic genome activation and DNA demethylation take place[68,69], and when transcriptional repression mediated by methylation of lysine 9 on histone H3 may have an important role[70]. Furthermore, recent integrants may be incompletely silenced[71]; if these proviruses affect gene expression, this can lead to phenotypic variation in the host[72]. Details of these complex processes, potentially involving hundreds of components, are beginning to emerge (reviewed in REF. 57) but lie outside the scope of this article.

Epigenetic mechanisms
Means by which gene expression can be modulated without altering the primary nucleotide sequences of the genes. Examples include cytosine methylation and histone deacetylation.

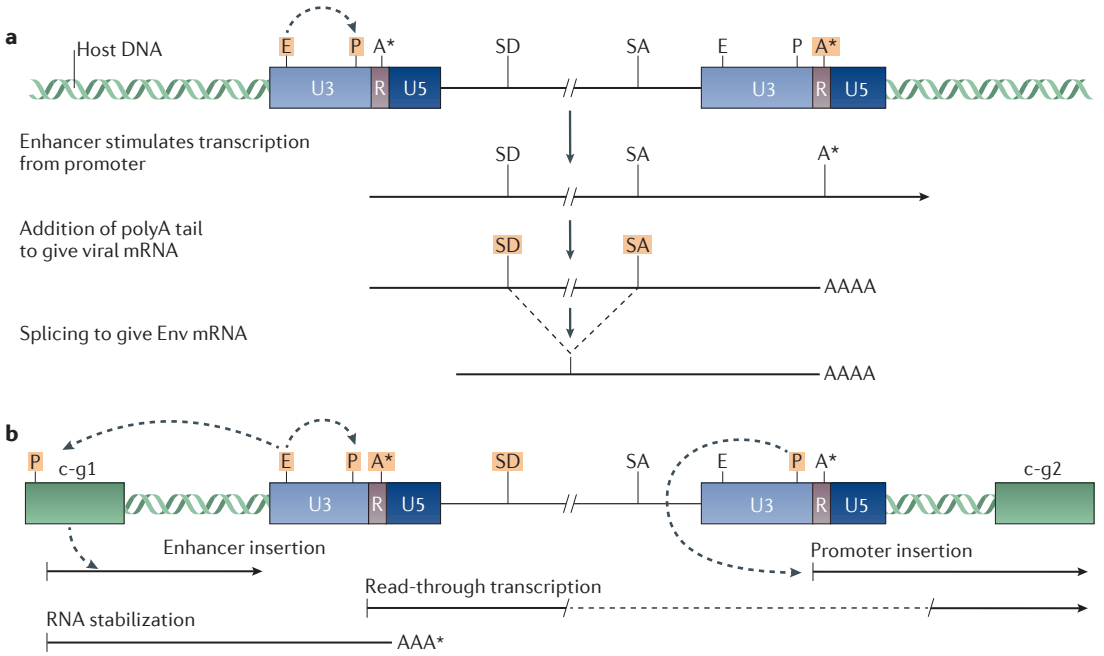

Figure 3 | **Different effects of retroviral regulatory sequences on viral and host RNA expression. a** | Effects on the virus. Within proviral DNA are found promoter (P), enhancer (E), polyadenylation (A*), splice donor (SD) and splice acceptor (SA) sequences (yellow highlighting indicates points at which these sequences are active). They are important for the control of viral RNA transcription, cleavage and polyadenylation, as well as for splicing, which gives rise to the Env-encoding mRNA. **b** | Effects on the host. The viral control sequences can also affect the transcription and stability of genes (referred to here as c-g1 and c-g2) neighbouring the integrated provirus by providing alternative enhancer, promoter and splicing elements, and resulting in read-through transcription or RNA stabilization. As discussed in the main text, this can result in tumorigenicity by activation of cellular oncogenes[156], mutation[157] or novel patterns of gene expression[43]. Not illustrated here are a number of important regulatory sequences present in retroviral genomes[3] that do not directly affect the expression of neighbouring sequences, such negative regulatory elements that can affect expression in embryonic cells, trans-acting regulatory sequences, sequences important for the export of unspliced RNA, and sequences for RNA dimerization and packaging.

Dedicated restriction factors. APOBEC3G was the first restriction factor identified that could block HIV-1 replication[73]. It belongs to a family of proteins containing one or two cytidine deaminase domains[74]. There are 11 members of this family in humans, several of which have antiviral activity[75]. A range of retroviruses, non-LTR retrotransposons and even a few viruses that do not contain reverse transcriptase are inhibited by APOBEC3G; however, its mechanism of action remains to be fully characterized. Initial studies showed that APOBEC3G incorporation into HIV-1 leads to hypermutation characterized by guanosine-to-adenosine transition mutations on plus-strand DNA, resulting from cytidine deamination on minus-strand DNA[76]. It was initially proposed that deamination would have a key role in restriction[77], but subsequent studies showed that deamination-deficient mutants could restrict virus replication[78]. Instead, it seems likely that APOBEC3G inhibits one or more steps of reverse transcription[79].

Another restriction factor, TRIM5α, seems to target the polymerized capsid protein, which surrounds the retroviral cores that enter cells following membrane fusion at cell surface or endosome[80]. It was originally identified by screening cells for resistance to HIV-1 following introduction of a rhesus macaque cDNA library[81]. TRIM5α from different species can restrict a range of retroviruses with specificity determined by binding to the carboxy-terminal B30.2 region of the protein[82–84]. Binding seems to be of low affinity and dependent on the common hexameric structure of the retroviral core, potential features of a mechanism designed to recognize multiple targets[85]. Binding can interfere with core stability — controlled uncoating is a key feature of the retroviral life cycle[86] — probably by a mechanism involving the proteasome[87]. Alternatively, it may result in sequestration of pre-integration complexes en route to the nucleus[88]. TRIM5α binding has also been reported to stimulate other cellular responses to retroviral infection[89].

Tetherin (also known as BST2), prevents the release of a range of enveloped viruses, including retroviruses, from the cell surface, resulting in chains of virions tethered to the membrane of producer cells[90,91]. Tetherin has a transmembrane anchor near its N terminus and a glycosyl phosphatidylinositol (GPI) anchor at its C terminus; in dimeric form, it seems to tether budded virions to the cells that they have just exited by crosslinking the cell and virion membranes[92].

Recently, the cellular factor SAM domain- and HD domain-containing 1 (SAMHD1) has been identified as an additional restriction factor[93,94]. However, transfer of

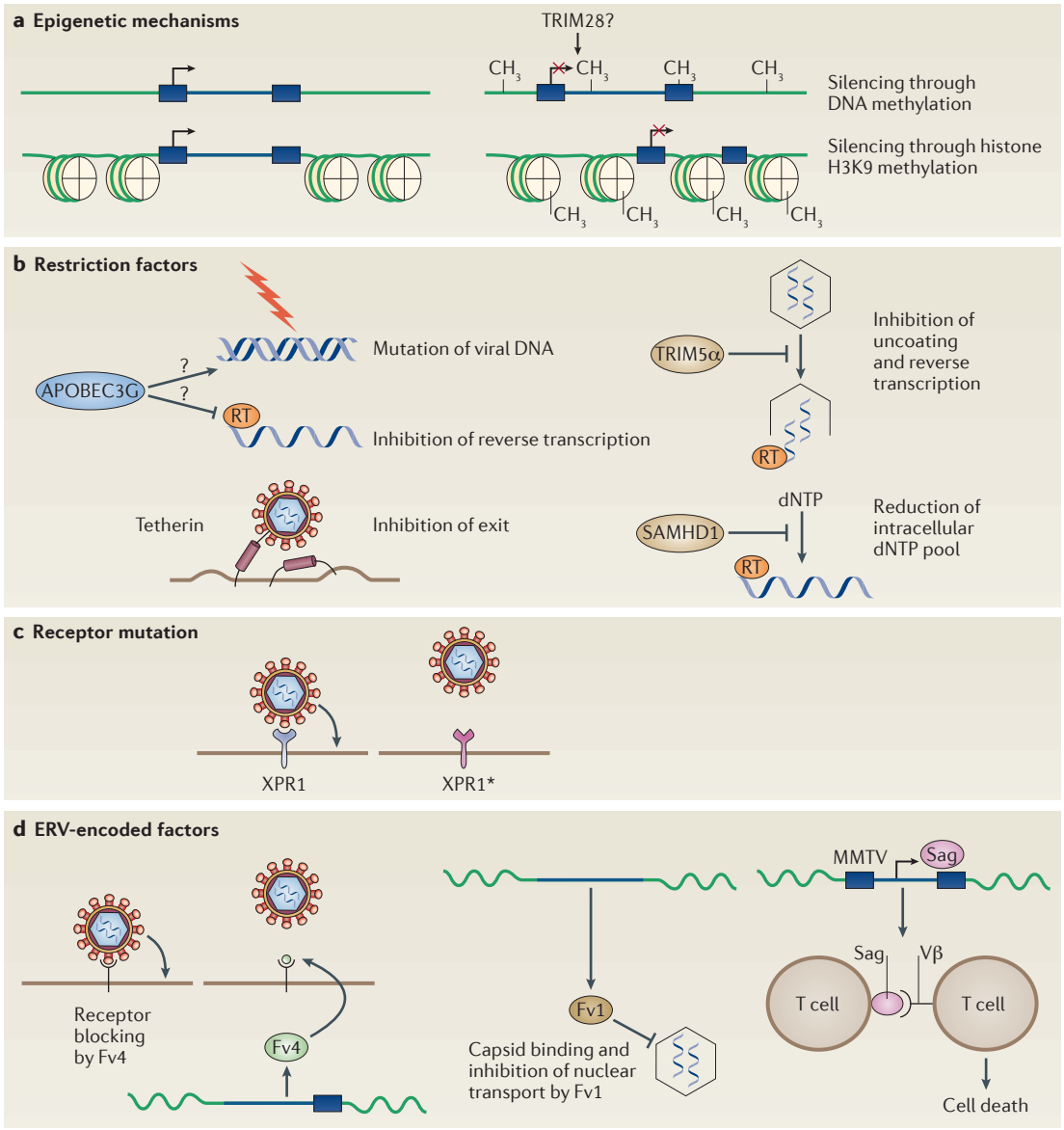

Figure 4 | **Different obstructions in the life cycle of retroviruses.** The viral life cycle can be blocked by the action of various factors that can be broadly divided into four main areas. **a** | Epigenetic mechanisms, including silencing by DNA methylation and histone methylation. **b** | Restriction factors, such as apolipoprotein B mRNA-editing complex 3 (APOBEC3G) and tripartite motif-containing 5α (TRIM5α), which inhibit reverse transcription, SAM domain- and HD domain-containing 1 (SAMHD1), which decreases cellular deoxyribonucleoside 5′-triphosphate (dNTP) levels, and tetherin, which prevents the release of viral progeny. The question mark denotes uncertainty about the mechanism of action. **c** | Receptor mutation, which prevents viral entry by mutation of entry receptors such as xenotropic and polytropic retrovirus receptor 1 (XPR1). **d** | Endogenous retrovirus (ERV)-encoded factors, including Friend virus susceptibility 4 (Fv4), which mediates masking or downregulation of a cellular receptor, Fv1 (MERV-L), which binds to the core of incoming retroviruses, and the Sag protein of endogenous mouse mammary tumour viruses (MMTVs), which stimulates specific T cell subsets and leads to cell death of target cells for exogenous MMTV. RT, reverse transcriptase.

SAMHD1 in most cell types does not seem to give rise to restriction of HIV-1. Furthermore, the demonstration that SAMDH1 has triphosphohydrolase activity raises the possibility that it acts to inhibit reverse transcription in certain cell types by limiting pools of deoxyribonucleoside 5′-triphosphates (dNTPs) rather than by a direct interaction with viruses[95].

APOBEC3G, TRIM5α and tetherin seem to be expressed constitutively and have thus been dubbed intrinsic restriction factors[96]. However, their expression, along with that of many other players in the field of innate immunity, is greatly enhanced after interferon treatment or induction[97]. Among such factors is three prime repair exonuclease 1 (TREX1)[98], which like SAMHD1 has been implicated in the human genetic disease Aicardi–Goutières syndrome[99]. Interestingly, TREX1 may contribute to the control of endogenous retroelement replication, as TREX1-knockout mice,

which usually die from an autoimmune condition, can be spared by treatment with reverse transcriptase-specific drugs[100]. It remains unknown how many other interferon-stimulated gene (ISG) products (for example, zinc-finger antiviral protein (ZAP), which prevents accumulation of cytoplasmic RNA from a range of viruses[101]) might contribute to the control of ERV movement. Nevertheless, it is clear that ERV acquisition occurs through infection of germ cells; thus, any factor acting to reduce viral load will reduce the probability of novel ERV formation.

Interacting and non-interacting cellular proteins. Receptor–co-receptor binding is a key step in retrovirus replication. Mutations that prevent binding will render the organism resistant to infection. Non-ecotropic MLV binding to its receptor, xenotropic and polytropic retrovirus receptor 1 (XPR1), on murine cells provides perhaps the best example of this resistance[60]. Apparently in response to germline colonization, a mutation in *Xpr1* has occurred in mice, resulting in the insensitivity of most inbred stains to a class of virus that is now labelled as xenotropic ('foreign-turning'). Other sub-species of mice (for example, *Mus castaneus*) carry an XPR1 protein that allows xenotropic MLV (XMV) binding and are thus susceptible to XMV infection. ERVs with xenotropic host ranges are found in several species, implying that this response is relatively common[102]. Other examples of resistance mediated by failure of viral proteins to bind specific cell factors are provided by studies of abortive HIV-1 replication in murine cells. At least three steps in the viral life cycle are blocked and, in each case, small differences between mouse and human proteins result in lack of binding[103,104]. Whether these changes result from selective pressures imposed by one or more unknown mouse lentiviruses is unknown.

The epigenetic mechanisms would seem to be the main control mechanism for ERVs. The other groups of factors would be expected to control ERVs that escape those blocks, as well as countering exogenous viruses that could give rise to novel ERVs or cause harm by other mechanisms.

Evolutionary interplay

Interactions between retroviruses and their hosts have had profound implications for human biology. Some changes, such as the use of cytosine methylation to control ERV expression, took place early in human evolution. Others, like the multiple assimilations of ERV *env* genes into placental function, are only slightly more recent. But the genetic interactions continue to this day, with the development and refinement of new host restriction factors and viral countermeasures associated with substantial genetic change on both sides. In light of the costs of the AIDS epidemic, I make no apologies for adopting a somewhat anthropomorphic approach to the discussion that follows on the genetic conflict between virus and host.

Complex retroviruses such as HIV-1 carry, in addition to the canonical retroviral genes *gag*, *pol* and *env*, a series of approximately six accessory genes[3]. Some, such as *tat* and *rev*, are involved in controlling viral RNA at the transcriptional and post-translational level[105], but others, such as *nef*, *vif* and *vpu*, seem to encode proteins designed to counter restriction factors, in particular ones that target features of retroviruses that cannot easily be changed, such as nucleic acids or membranes[58,106]. For example, Vif, which is present in most but not all lentiviruses[23], is an APOBEC3G-specific factor[74]. Vif induces polyubiquitylation and subsequent proteasomal degradation of APOBEC3G in a species-specific manner[107] through recruitment of a cellular ubiquitin ligase complex[108], thereby preventing APOBEC3G incorporation into virions. Sometimes different viruses use different accessory proteins for the same purpose. Thus, tetherin-mediated restriction can be overcome in different ways in a virus-specific manner: HIV-1 and simian immunodeficiency virus (SIV) use the accessory proteins Vpu and Nef, respectively, whereas in HIV-2 this countermeasure is a property of the Env protein[109,110]. Alternatively, one lentivirus may use an accessory factor to overcome a particular cellular factor, whereas another may not have an equivalent activity, as is the case with SAMHD1, which HIV-2 can block with Vpx but HIV-1 cannot[111]. In this respect, it should be emphasized that simple retroviruses without accessory genes also seem to have evolved means of circumventing restriction factors, for example by replicating in cell types with low levels of factor expression or by excluding factors from virions.

When confronted with a novel viral threat, it seems likely that hosts will first seek to modify and redeploy their existing antiviral weapons. A good example of a protein that is subject to continuous positive selection, presumably resulting from continual exposure to new viruses, is provided by TRIM5α[112]. Among other evolutionary changes, the *TRIM5* locus has undergone gene duplication and loss[113,114], high levels of non-synonymous amino-acid substitution[115] and duplication[116], and at least two cases in which the exon encoding the viral binding domain has been replaced by cyclophilin A, presumably following LINE-mediated retrotransposition[117,118]. Rapid evolution of APOBEC3, tetherin and SAMHD1 has also occurred[119–121]. Positively selected non-synonymous changes will tend to occur at, and can be used to identify, sites of interaction between two proteins involved in genetic conflict[115]. Interestingly, the ability to restrict a particular virus by different factors can be lost during evolution as well as gained[122,123]. Such changes may reflect key determinants of retroviral specificity[124,125].

Alternatively, the host may sometimes develop alternative defensive strategies, perhaps by enlisting ERVs (FIG. 4d). One recurring example, typified by the mouse gene Friend virus susceptibility 4 (*Fv4*)[126], is that of ERV-encoded Env proteins that bind viral receptors, thereby blocking infection by exogenous MLVs, in a manner analogous to the property of superinfection resistance shown by replicating retroviruses (FIG. 4d). ERVs in at least three different species of mice[127] and in some chickens[128] and cats[129] seem to operate in an identical fashion. A second example is *Fv1*, one of the first genes to be shown to affect host sensitivity to retrovirus infection[130]. *Fv1* encodes a protein of around 50 kDa that, like TRIM5α,

binds to the assembled cores of incoming retroviruses[131] and interferes with a post-penetration, pre-integration step in the retroviral life cycle (FIG. 4d). *Fv1* is derived from the *gag* gene of an ERV related to mouse ERV-L (MERV-L)[132]. The capsid-binding restriction factors are a further example of nature reusing a successful strategy — factors targeting the assembled retroviral core have evolved on at least four occasions[118]. A third case of ERVs conferring resistance to exogenous retrovirus infection is provided by the MMTVs. The MMTV *sag* gene product is a superantigen, capable of specific binding to the *Vb* gene product of murine T cells[133]. In the context of an exogenous virus infection, this can stimulate cell proliferation and facilitate infection by increasing the number of cell targets, but Sag expression from an endogenous MMTV will lead to clonal T-cell deletion and resistance to infection owing to the absence of targets[134] (FIG. 4d).

For the most part, restriction factors seem to present a formidable barrier to cross-species transfer of infectious viruses and may provide a substantial driving force for lentiviral evolution[135]. However, the existence of ERVs provides evidence that such barriers have been breached on multiple occasions. Unless restriction factors are absolutely effective, they will sometimes allow cross-species transfer to occur. In some cases (for example, that of group P HIV-1 infection of humans[136]), virus replication may be self-limiting and little or no onward spread in the new species occurs. However, virus adaptation might occur, leading to higher levels of replication and spread[137]. Indeed, indirect evidence suggests that low-level restriction factor expression might itself be mutagenic, providing a mechanism for viral genetic change[138]. However, the scope for mutagenic change in viral genes must be limited by functional considerations[139,140]. Alternatively, novel or re-targeted accessory proteins might be developed. For example, Vpx is clearly related to Vpr[141]; it seems reasonable to assume that a gene duplication event followed by a sequence change gave rise to the SAMHD1-specific activity characteristic of HIV-2 and certain SIVs[142]. Similarly, the evolution of tetherin has moulded the functions of lentiviral Nef and Vpu[119].

An extreme response would be for the virus to change its replication strategy to avoid extracellular restriction factors or the need to bind cellular receptors. Such changes have apparently happened on at least two occasions. Two groups of murine ERVs — intracisternal A-particles (IAPs) and MusD elements — have developed a replication strategy lacking an extracellular phase. In both cases, *env* has been deleted and a mutation in Gag allows intracellular assembly and budding[143]. Such ERVs have clearly undergone a different evolutionary pathway from retrotransposons such as Ty1 and Ty3, which probably never had *env* genes. This can be seen as an extreme case of molecular parasitism, with loss of the ability to 'move house' and thus absolute sensitivity to intracellular restriction factors such as TREX1 (REF. 98).

Concluding remarks

It is now abundantly clear that retroviruses and their hosts have coexisted for tens of millions of years and that this has led to substantial genetic change on both sides. One apparent consequence has been the development of retrovirus-specific restriction factors that together provide defence against retrovirus infection or mobilization in most cases. A degree of redundancy is presumably important to control a more 'genetically agile' foe[125]. However, it remains unclear how these factors work on a population level. Two interrelated questions that perhaps merit further investigation are whether the main targets for restriction factors are endogenous or exogenous retroviruses, and which class of element is the primary driver for evolutionary change.

There is a paucity of experimental systems that will allow a direct comparison of the threats posed, or the selective pressures exerted, by similar endogenous and exogenous viruses. A further complication is evaluating the relative threat posed by retroviruses causing diseases with different pathologies and time courses. For example, how does one compare the threat posed by agents that operate at different speeds? Or by agents that act by non-specific genome bombardment with those requiring integration in a specific location?

It is tempting to speculate that restriction factors have evolved to prevent ERV amplification or disease associated with their expression. Restriction factors can clearly protect against the consequences of infection by either endogenous or exogenous viruses by interfering with virus replication. For example, although *Fv1* was originally described as a cellular gene conferring susceptibility to MLVs such as Friend virus, it can also protect AKR mice from spontaneous thymomas associated with the activation of endogenous MLVs[144]. Furthermore, ERVs bearing the signatures of APOBEC3-mediated cytidine deamination have been described[20,145]. The continuous presence of ERVs might impart a greater selective force than occasional exposure to exogenous viruses. Bursts of ERV amplification have been reported[146,147]. It will be of considerable interest to generate replication-competent viruses, or viruses carrying the protein targets for restriction factors[140], associated with these bursts of ERV amplification and test for sensitivity to re-synthesized restriction factors from before or after the time of bursting.

However, it is unclear how great are the selective pressures imposed by, for example, the disease characterizing AKR mice, as cancers have not been found in young animals. Thus, it is unclear what effect such a disease would have on animal populations in the wild. Moreover, it seems that in the apparent absence of endogenous lentiviruses[23], selection of *TRIM–CYP* fusion genes has occurred independently in both owl monkeys and macaques. Selection of these factors must therefore be driven by exogenous viruses, which are presumably present in a high proportion of the individuals in which selection is taking place. It is tempting to speculate that HIV-1, if left unchecked, might also eventually select a similar restriction factor. In any event, restriction factors have proved to be effective natural antiviral agents, and it remains a tantalizing possibility that understanding their mode of action might facilitate the development of novel antiviral strategies.

Note added in proof

The evolutionary success of ERVs that have lost their *env* genes is illustrated well by a very recent study[158]. Analysis of ERV loci from 38 mammalian species by *in silico* screening showed that *env* gene loss is associated with a significant (up to 30-fold) increase in genome copy number, implying that such *env*-less ERVs can be considered genomic superspreaders.

1. Coffin, J. M., Hughes, S. H. & Varmus, H. E. (eds) *Retroviruses* (Cold Spring Harbor Press, 1997).
2. Kurth, R. & Bannert, N. (eds) *Retroviruses: Molecular Biology, Genomics and Pathogenesis* (eds Kurth, R. & Bannert, N.) (Caister Academic, 2010).
3. Goff, S. P. in *Fields Virology* Ch. 55 (eds Knipe, D. M. *et al.*) 1999–2069 (Lippincott Williams & Wilkins, 2007).
4. Hirsch, A. J. The use of RNAi-based screens to identify host proteins involved in viral replication. *Future Microbiol.* **5**, 303–311 (2010).
5. Boeke, J. D. & Stoye, J. P. Retrotransposons, endogenous retroviruses, and the evolution of retroelements. *Retroviruses* [online], http://www.ncbi.nlm.nih.gov/books/NBK19468 (1997).
A good introduction to the basic properties of ERVs.
6. Weiss, R. A. The discovery of endogenous retroviruses. *Retrovirology* **3**, 67 (2006).
A personal account of the history of ERV discovery by one of the main players in the field.
7. Rosenberg, N. & Jolicoeur, P. in *Retroviruses* (eds Coffin, J. M., Hughes, S. H. & Varmus, H. E.) 475–585 (Cold Spring Harbor Press, 1997).
8. Martin, M. A., Bryan, T., Rasheed, S. & Khan, A. S. Identification and cloning of endogenous retroviral sequences present in human DNA. *Proc. Natl Acad. Sci. USA* **78**, 4892–4896 (1981).
9. Medstrand, P. & Blomberg, J. Characterization of novel reverse transcriptase encoding human endigenous retroviral sequences similar to type A and type B retroviruses: differential transcription in normal human tissues. *J. Virol.* **67**, 6778–6787 (1993).
10. McAllister, R. M. *et al.* C-type virus released from cultured human rhabdomyosarcoma cells. *Nature New Biol.* **235**, 3–6 (1972).
11. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
12. Mouse Genome Sequencing Consortium *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
13. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
14. Lerat, E. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. *Heredity* **104**, 520–522 (2010).
15. Tristem, M. Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. *J. Virol.* **74**, 3715–3730 (2000).
One of the first detailed studies analysing the diversity of human ERVs.
16. Katzourakis, A. & Tristem, M. in *Retroviruses and Primate Genome Evolution* (ed. Sverdlov, E. D.) 186–203 (Landes Bioscience, 2005).
17. Stocking, C. & Kozak, C. A. Murine endogenous retroviruses. *Cell. Mol. Life Sci.* **65**, 3383–3398 (2008).
18. Huda, A., Polavarapu, N., Jordan, I. K. & McDonald, J. F. Endogenous retroviruses of the chicken genome. *Biol. Direct* **3**, 9 (2008).
19. Blomberg, J., Benachenhou, F., Blikstad, V., Sperber, G. & Mayer, J. Classification and nomenclature of endogenous retroviral sequences (ERVs): problems and recommendations. *Gene* **448**, 115–123 (2009).
20. Jern, P., Stoye, J. P. & Coffin, J. M. Role of APOBEC3 in genetic diversity among endogenous murine leukemia viruses. *PLoS Genet.* **3**, e183 (2007).
21. Benachenhou, F. *et al.* Evolutionary conservation of orthoretroviral long terminal repeats (LTRs) and ab initio detection of single LTRs in genomic data. *PLoS ONE* **4**, e5179 (2009).
22. Copeland, N. G., Hutchison, K. W. & Jenkins, N. A. Excision of the DBA ecotropic provirus in dilute coat-color revertants of mice occurs by homologous recombination involving the viral LTRs. *Cell* **33**, 379–387 (1983).
23. Katzourakis, A., Tristem, M., Pybus, O. G. & Gifford, R. J. Discovery and analysis of the first endogenous lentivirus. *Proc. Natl Acad. Sci. USA* **104**, 6261–6265 (2007).
24. Gilbert, C., Maxfield, D. G., Goodman, S. M. & Feschotte, C. Parallel germline infiltration of a lentivirus in two Malagasy lemurs. *PLoS Genet.* **5**, e10000425 (2009).
25. Katzourakis, A., Gifford, R. J., Tristem, M., Gilbert, M. T. P. & Pybus, O. G. Macroevolution of complex retroviruses. *Science* **325**, 1512 (2009).
26. Polavarapu, N., Bowen, N. J. & McDonald, J. F. Identification, characterization and comparative genomics of chimpanzee endogenous retroviruses. *Genome Biol.* **7**, R51 (2006).
27. Smit, A. F. A. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**, 657–663 (1999).
28. Goodier, J. L. & Kazazian, H. H. Jr. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* **135**, 23–35 (2008).
29. Holmes, E. C. The evolution of endogenous viral elements. *Cell Host Microbe* **10**, 368–377 (2011).
A review considering the evolutionary implications of endogenous viral sequences of non-retroviral origin.
30. Tarlington, R. E., Meers, J. & Young, P. R. Retroviral invasion of the koala genome. *Nature* **442**, 79–81 (2006).
31. Jaenisch, R. Germ line integration and endogenous transmission of the exogenous Moloney leukemia virus. *Proc. Natl Acad. Sci. USA* **73**, 1260–1264 (1976).
32. Salter, D. W., Smith, E. J., Hughes, S. H., Wright, S. E. & Crittenden, L. B. Transgenic chickens: insertion of retroviral genes into the chicken germ line. *Virology* **157**, 236–240 (1987).
33. Lock, L. F., Keshet, E., Gilbert, D. J., Jenkins, N. A. & Copeland, N. G. Studies on the mechanism of spontaneous germline ecotropic provirus acquisition in mice. *EMBO J.* **7**, 4169–4177 (1988).
34. Belshaw, R. *et al.* Long-term reinfection of the human genome by endogenous retroviruses. *Proc. Natl Acad. Sci. USA* **101**, 4894–4899 (2004).
35. Nakagawa, K. & Harrison, L. C. The potential roles of endogenous retroviruses in autoimmunity. *Immunol. Rev.* **152**, 193–236 (1996).
36. Ruprecht, K., Mayer, J., Sauter, M., Roemer, K. & Mueller-Lantzsch, N. Endogenous retroviruses and cancer. *Cell. Mol. Life Sci.* **65**, 3366–3382 (2008).
37. Voisset, C., Weiss, R. A. & Griffiths, D. J. Human RNA "rumor" viruses: the search for novel human retroviruses in chronic disease. *Microbiol. Mol. Biol. Rev.* **72**, 157–196 (2008).
An at times amusing account of the problems and pitfalls encountered in trying to identify possible retroviral causes for human diseases.
38. Li, F., Nellaker, C., Yolken, R. H. & Karlsson, H. A systematic evaluation of expression of HERV-W elements; influence of genomic context, viral structure and orientation. *BMC Genomics* **12**, 22 (2011).
39. Kaufmann, S. *et al.* Human endogenous retrovirus protein Rec. interacts with the testicular zinc-finger protein and androgen receptor. *J. Gen. Virol.* **91**, 1494–1502 (2010).
40. Lauring, A. S., Anderson, M. M. & Overbaugh, J. Specificity in receptor usage by T-cell-tropic feline leukemia viruses: implications for the *in vivo* tropism of immunodeficiency-inducing variants. *J. Virol.* **75**, 8888–8898 (2001).
41. Stoye, J. P., Moroni, C. & Coffin, J. M. Virological events leading to spontaneous AKR thymomas. *J. Virol.* **65**, 1273–1285 (1991).
42. Maksakova, I. A. *et al.* Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet.* **2**, e2 (2006).
43. Cohen, C. J., Lock, W. M. & Mager, D. L. Endogenous retroviral LTRs as promoters for human genes: a critical assessment. *Gene* **448**, 105–114 (2009).
44. Lamprecht, B., Bonifer, C. & Mathas, S. Repeat-element driven activation of proto-oncogenes in human malignancies. *Cell Cycle* **9**, 4276–4281 (2010).
45. Zhang, Y., Romanish, M. T. & Mager, D. L. Distributions of transposable elements reveal hazardous zones in mammalian introns. *PLoS Comput. Biol.* **7**, e1002046 (2011).
46. Hughes, J. F. & Coffin, J. M. Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. *Nature Genet.* **29**, 487–489 (2001).
47. Romanish, M. T., Cohen, C. J. & Mager, D. L. Potential mechanisms of endogenous retroviral-mediated genomic instability in human cancer. *Semin. Cancer Biol.* **20**, 246–253 (2010).
48. Sun, C. *et al.* Deletion of the *azoospermia factor a* (*AZFa*) region of human Y chromosome caused by recombination between HERV15 proviruses. *Hum. Mol. Genet.* **9**, 2391–2396 (2000).
49. Mi, S. *et al.* Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785–789 (2000).
50. Dupressoir, A. *et al.* Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc. Natl Acad. Sci. USA* **106**, 12127–12132 (2009).
The first unequivocal evidence that a protein encoded by an ERV has a key role in vertebrate development.
51. Mangeney, M. *et al.* Placental syncytins: genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. *Proc. Natl Acad. Sci. USA* **104**, 20534–20539 (2007).
52. Heidmann, O., Vernochet, C., Dupressoir, A. & Heidmann, T. Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new "syncytin" in a third order of mammals. *Retrovirology* **6**, 107 (2009).
53. Meisler, M. H. & Ting, C. N. The remarkable evolutionary history of the human amylase genes. *Crit. Rev. Oral Biol. Med.* **4**, 503–509 (1993).
54. Beyer, U., Moll-Rocek, J., Moll, U. M. & Dobbelstein, M. Endogenous retrovirus drives hitherto unknown proapoptotic p63 isoforms in the male germ line of humans and great apes. *Proc. Natl Acad. Sci. USA* **108**, 3624–3629 (2011).
55. Walsh, C. P. & Bestor, T. H. Cytosine methylation and mammalian development. *Genes Dev.* **13**, 26–34 (1999).
56. Maksakova, I. A., Mager, D. L. & Reiss, D. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cell. Mol. Life Sci.* **65**, 3329–3347 (2008).
57. Rowe, H. M. & Trono, D. Dynamic control of endogenous retroviruses during development. *Virology* **411**, 273–287 (2011).
58. Neil, S. & Bieniasz, P. Human immunodeficiency virus, restriction factors, and interferon. *J. Interferon Cytokine Res.* **29**, 569–580 (2009).
59. Bushman, F. D. *et al.* Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog.* **5**, e1000437 (2009).
60. Kozak, C. A. The mouse "xenotropic" gammaretroviruses and their XPR1 receptor. *Retrovirology* **7**, 101 (2010).
A comprehensive review of the evolutionary changes occurring in the mouse receptor for the XMV.
61. Groudine, M., Eisenman, R. & Weintraub, H. Chromatin structure of endogenous retroviral genomes and activation by an inhibitor of DNA methylation. *Nature* **292**, 311–317 (1981).
62. Jähner, D. *et al. De novo* methylation and expression of retroviral genomes during mouse embryogenesis. *Nature* **298**, 623–628 (1982).
63. Wigler, M., Levy, D. & Perucho, M. The somatic replication of DNA methylation. *Cell* **24**, 33–40 (1981).
64. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335–340 (1997).
65. Wolf, D. & Goff, S. P. Embyonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **458**, 1201–1204 (2009).
66. Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 (2010).

67. Nisole, S., Stoye, J. P. & Saïb, A. Trim family proteins: retroviral restriction and antiviral defence. *Nature Rev. Microbiol.* **3**, 799–808 (2005).
68. Reik, W. Stability and flexibility of epigentic gene regulation in mammalian development. *Nature* **411**, 425–432 (2007).
69. Feng, S., Jacobsen, S. E. & Reik, W. Epigenetic reprogramming in plant and animal development. *Science* **330**, 622–627 (2010).
70. Leung, D. C. & Lorincz, M. C. Silencing of endogenous retroviruses: when and why do histone marks predominate? *Trends Biochem. Sci.* **37**, 127–133 (2012).
71. Reiss, D. & Mager, D. L. Stochastic epigenetic silencing of retrotransposons: does stability come with age? *Gene* **390**, 130–135 (2007).
72. Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nature Genet.* **23**, 314–318 (1999).
73. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646–650 (2002).
The first report of the cloning of the host resistance gene being overcome by the lentivirus accessory gene *vif*.
74. Malim, M. H. APOBEC proteins and intrinsic resistance to HIV-1 infection. *Phil. Trans. R. Soc. B* **364**, 675–687 (2009).
75. Chiu, Y. L. & Greene, W. C. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu. Rev. Immunol.* **26**, 317–353 (2008).
76. Harris, R. S. *et al.* DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803–809 (2003).
77. Harris, R. S., Sheehy, A. M., Craig, H. M., Malim, M. H. & Neuberger, M. S. DNA deamination: not just a trigger for antibody diversification but also a mechanism for defense against retroviruses. *Nature Immunol.* **4**, 641–643 (2003).
78. Newman, E. N. *et al.* Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr. Biol.* **15**, 166–170 (2005).
79. Mbisa, J. L., Bu, W. & Pathak, V. K. APOBEC3F and APOBEC3G inhibit HIV-1 DNA integration by different mechanisms. *J. Virol.* **84**, 5250–5259 (2010).
80. Huthoff, H. & Towers, G. J. Restriction of retroviral replication by APOBEC3G/F and TRIM5α. *Trends Microbiol.* **16**, 612–619 (2008).
81. Stremlau, M. *et al.* The cytoplasmic body component TRIM5α restricts HIV-1 infection in Old World monkeys. *Nature* **427**, 848–853 (2004).
The initial description of TRIM5α as a cellular factor inhibiting HIV-1 replication.
82. Stremlau, M. *et al.* Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5α restriction factor. *Proc. Natl Acad. Sci. USA* **103**, 5514–5519 (2006).
83. Perez-Caballero, D., Hatziioannou, T., Yang, A., Cowan, S. & Bieniasz, P. D. Human tripartite motif 5α domains responsible for retrovirus restriction activity and specificity. *J. Virol.* **79**, 8969–8978 (2005).
84. Ohkura, S., Yap, M. W., Sheldon, T. & Stoye, J. P. All three variable regions of the TRIM5α B30.2 domain can contribute to the specificity of retrovirus restriction. *J. Virol.* **80**, 8554–8565 (2006).
85. Ganser-Pornillos, B. K. *et al.* Hexagonal assembly of a restricting TRIM5α protein. *Proc. Natl Acad. Sci. USA* **108**, 534–539 (2011).
86. Forshey, B. M., von Schwedler, U., Sundquist, W. I. & Aiken, S. C. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J. Virol.* **76**, 5667–5677 (2002).
87. Wu, X., Anderson, J. L., Campbell, E. M., Joseph, A. M. & Hope, T. J. Proteasome inhibitors uncouple rhesus TRIM5α restriction of HIV-1 reverse transcription and infection. *Proc. Natl Acad. Sci. USA* **103**, 7465–7470 (2006).
88. Campbell, E. M., Perez, O., Anderson, J. L. & Hope, T. J. Visualization of a proteasome-independent intermediate during restriction of HIV-1 by rhesus TRIM5α. *J. Cell Biol.* **180**, 549–561 (2008).
89. Pertel, T. *et al.* TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* **472**, 361–365 (2011).
90. Neil, S. J. D., Zang, T. & Bieniasz, P. D. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451**, 425–430 (2008).
A paper providing the first description of tetherin as a factor preventing retrovirus release from the cell surface.
91. Martin-Serrano, J. & Neil, S. J. Host factors involved in retroviral budding and release. *Nature Rev. Microbiol.* **9**, 519–531 (2011).
92. Evans, D. T., Serra-Moreno, R., Singh, R. K. & Guatelli, J. C. BST-2/tetherin: a new component of the innate immune response to enveloped viruses. *Trends Microbiol.* **18**, 388–396 (2010).
93. Laguette, N. *et al.* SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* **474**, 654–657 (2011).
94. Hrecka, K. *et al.* Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* **474**, 658–661 (2011).
95. Goldstone, D. C. *et al.* HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature* **480**, 379–382 (2011).
96. Bieniasz, P. D. Intrinsic immunity: a front-line defense against viral attack. *Nature Immun.* **5**, 1109–1115 (2004).
97. Schoggins, J. W. & Rice, C. M. Interferon-stimulated genes and their antiviral receptor function. *Curr. Opin. Virol.* **1**, 1–7 (2011).
98. Stetson, D. B., Ko, J. S., Heidmann, T. & Medzhitov, R. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* **134**, 587–598 (2008).
99. Rice, G. I. *et al.* Mutations involved in Aicardi–Goutières syndrome implicate SAMHD1 as regulator of the innate immune response. *Nature Genet.* **41**, 829–832 (2009).
100. Beck-Engeser, G. B., Eilat, D. & Wabl, M. An autoimmune disease prevented by anti-retroviral drugs. *Retrovirology* **8**, 91 (2011).
101. Gao, G., Guo, X. & Goff, S. P. Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. *Science* **297**, 1703–1706 (2002).
102. Dewannieux, M., Ribet, D. & Heidmann, T. Risks linked to endogenous retroviruses for vaccine production: a general overview. *Biologicals* **38**, 366–370 (2010).
103. Bieniasz, P. D. & Cullen, B. R. Multiple blocks to human immunodeficiency virus type 1 replication in rodent cells. *J. Virol.* **74**, 9868–9877 (2000).
104. Sherer, N. M. *et al.* Evolution of a species-specific determinant within human CRM1 that regulates the post-transcriptional phases of HIV-1 replication. *PLoS Pathog.* **7**, e1002395 (2011).
105. Cullen, B. R. Mechanism of action of regulatory proteins encoded by complex retroviruses. *Microbiol. Rev.* **56**, 375–394 (1992).
106. Malim, M. H. & Emerman, M. HIV-1 accessory proteins—ensuring viral survival in a hostile environment. *Cell Host Microbe* **3**, 388–398 (2008).
A penetrating review considering the role of viral accessory factors in overcoming host restriction factors.
107. Simon, J. H. *et al.* The regulation of primate immunodeficiency virus infectivity by Vif is cell species restricted: a role for Vif in determining virus host range and cross-species transmission. *EMBO J.* **17**, 1259–1267 (1998).
108. Yu, X. *et al.* Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif–Cul5–SCF complex. *Science* **302**, 1056–1060 (2003).
109. Douglas, J. L. *et al.* The great escape: viral strategies to counter BST-2/tetherin. *PLoS Pathog.* **6**, e1000913 (2010).
110. Le Tortorec, A., Willey, S. & Neil, S. J. Antiviral inhibition of enveloped virus release by tetherin/BST-2: action and counteraction. *Viruses* **3**, 520–540 (2011).
111. Kaushik, R., Zhu, X., Stranska, R., Wu, Y. & Stevenson, M. A cellular restriction dictates the permissivity of nondividing monocytes/macrophages to lentivirus and gammaretrovirus infection. *Cell Host Microbe* **6**, 68–80 (2009).
112. Johnson, W. E. & Sawyer, S. L. Molecular evolution of the antiretroviral *TRIM5* gene. *Immunogenetics* **61**, 163–178 (2009).
A stimulating review considering the role of positive selection in modulating the evolution of the *TRIM5* gene.
113. Ylinen, L. M. J. *et al.* Isolation of an active *Lv1* gene from cattle indicates that tripartite motif protein-mediated innate immunity to retroviral infection is widespread among mammals. *J. Virol.* **80**, 7332–7338 (2006).
114. Tareen, S. U., Sawyer, S. L., Malik, H. S. & Emerman, M. An expanded clade of rodent *Trim5* genes. *Virology* **385**, 473–483 (2009).
115. Sawyer, S. L., Wu, L. I., Emerman, M. & Malik, H. S. Positive selection of primate TRIM5α identifies a critical species-specific retroviral restriction domain. *Proc. Natl Acad. Sci. USA* **102**, 2832–2837 (2005).
116. Song, B. *et al.* The B30.2(SPRY) domain of retroviral restriction factor TRIM5α exhibits lineage-specific length and sequence variation in primates. *J. Virol.* **79**, 6111–6121 (2005).
117. Sayah, D. M., Sokolskaja, E., Berthoux, L. & Luban, J. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* **430**, 569–573 (2004).
118. Stoye, J. P. & Yap, M. W. Chance favors a prepared genome. *Proc. Natl Acad. Sci. USA* **105**, 3177–3178 (2008).
Describes a series of papers providing evidence that the evolution of restriction factors is continuing.
119. Lim, E. S., Malik, H. S. & Emerman, M. Ancient adaptive evolution of tetherin shaped the functions of Vpu and Nef in human immunodeficiency virus and primate lentiviruses. *J. Virol.* **84**, 7124–7134 (2010).
120. McNatt, M. W. *et al.* Species-specific activity of HIV-1 Vpu and positive selection of tetherin transmembrane domain variants. *PLoS Pathog.* **5**, e1000300 (2009).
121. Planelles, V. SAMHD1 joins the Red Queen's court. *Cell Host Microbe* **16**, 103–105 (2012).
122. Goldschmidt, V. *et al.* Antiretroviral activity of ancestral TRIM5α. *J. Virol.* **82**, 2089–2096 (2008).
123. OhAinle, M., Kerns, J. A., Li, M. M., Malik, H. S. & Emerman, M. Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe* **4**, 249–259 (2008).
124. Yap, M. W., Nisole, S. & Stoye, J. P. A single amino acid change in the SPRY domain of human TRIM5α leads to HIV-1 restriction. *Curr. Biol.* **15**, 73–78 (2005).
125. Meyerson, N. R. & Sawyer, S. L. Two-stepping through time: mammals and viruses. *Trends Micribiol.* **19**, 286–294 (2011).
126. Odaka, T., Ikeda, H. & Akatsuka, T. Restricted expression of endogenous N-tropic XC-positive leukemia virus in hybrids between G and AKR mice: an effect of the *Fv-4r* gene. *Int. J. Cancer* **25**, 757–762 (1980).
127. Wu, T., Yan, Y. & A., K. C. *Rmcf2*, a xenotropic provirus in the Asian mouse species *Mus castaneus*, blocks infection by mouse gammaretroviruses. *J. Virol.* **79**, 9677–9684 (2005).
128. Robinson, H. L. & Lamoreux, W. F. Expression of endogenous ALV antigens and susceptibility to subgroup E ALV in three strains of chickens (endogenous avian C-type virus). *Virology* **69**, 50–62 (1976).
129. McDougall, A. S. *et al.* Defective endogenous proviruses are expressed in feline lymphoid cells: evidence for a role in natural resistance to subgroup B feline leukemia virus. *J. Virol.* **68**, 2151–2160 (1994).
130. Lilly, F. *Fv-2*: Identification and location of a second gene governing the spleen focus response to Friend leukemia virus in mice. *J. Natl Cancer Inst.* **45**, 163–169 (1970).
131. Hilditch, L. *et al.* Ordered assembly of murine leukemia virus capsid protein on lipid nanotubes directs specific binding by the restriction factor, Fv1. *Proc. Natl Acad. Sci. USA* **108**, 5771–5776 (2011).
132. Best, S., Le Tissier, P., Towers, G. & Stoye, J. P. Positional cloning of the mouse retrovirus restriction gene *Fv1*. *Nature* **382**, 826–829 (1996).
The first cloning of a restriction factor revealed how ERVs could become antiviral factors.
133. Czarneski, J., Rassa, J. C. & Ross, S. R. Mouse mammary tumor virus and the immune system. *Immunol. Res.* **27**, 469–480 (2003).
134. Frankel, W. N., Rudy, C., Coffin, J. M. & Huber, B. T. Linkage of *Mls* genes to endogenous mammary tumour viruses of inbred mice. *Nature* **349**, 526–528 (1991).
135. Gifford, R. J. Viral evolution in deep time: lentiviruses and mammals. *Trends Genet.* **28**, 89–100 (2012).
136. Plantier, J. C. *et al.* A new human immunodeficiency virus derived from gorillas. *Nature Med.* **15**, 871–872 (2009).
137. Kirmaier, A. *et al.* 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 (2010).
Provides evidence that restriction factors act to suppress cross-species transmission and can drive virus evolution.
138. Kim, E. Y. *et al.* Human APOBEC3G-mediated editing can promote HIV-1 sequence diversification and accelerate adaptation to selective pressure. *J. Virol.* **84**, 10402–10405 (2010).

139. Rein, A. Genetic fingerprinting of a retroviral *gag* gene suggests an important role in virus replication. *Proc. Natl Acad. Sci. USA* **100**, 11929–11930 (2003).
140. Goldstone, D. C. *et al.* Structural and functional analysis of prehistoric lentiviruses uncovers an ancient molecular interface. *Cell Host Microbe* **8**, 248–259 (2010).
141. Tristem, M., Marshall, C., Karpas, A. & Hill, F. Evolution of the primate lentivirus: evidence from vpx and vpr. *EMBO J.* **11**, 3405–3412 (1992).
142. Sharova, N. *et al.* Primate lentiviral Vpx commandeers DDB1 to counteract a macrophage restriction. *PLoS Pathog.* **4**, e1000057 (2008).
143. Ribet, D. *et al.* An infectious progenitor for the murine IAP retrotransposon: emergence of an intracellular genetic parasite from an an ancient retrovirus. *Genome Res.* **18**, 597–609 (2008).
144. Haran-Ghera, N., Peled, A., Brightman, B. K. & Fan, H. Lymphomagenesis in AKR.*Fv-1*^{b} congenic mice. *Cancer Res.* **53**, 3433–3438 (1993).
145. Lee, Y. N., Malim, M. H. & Bieniasz, P. D. Hypermutation of an ancient human retrovirus by APOBEC3G. *J. Virol.* **82**, 8762–8770 (2008).
146. Goodchild, N. L., Wilkinson, D. A. & Mager, D. L. Recent evolutionary expansion of a subfamily of RTVL-H human endogenous retrovirus-like elements. *Virology* **196**, 778–788 (1993).
147. Cordonnier, A., Casella, J.-F. & Heidmann, T. Isolation of novel human endogenous retrovirus-like elements with foamy virus-related *pol* sequence. *J. Virol.* **69**, 5890–5897 (1995).
148. Johnson, W. E. & Coffin, J. M. Constructing primate phylogenies from ancient retrovirus sequences. *Proc. Natl Acad. Sci. USA* **96**, 10254–10260 (1999).
149. Martins, H. & Villesen, P. Improved integration time estimation of endogenous retroviruses with phylogenetic data. *PLoS ONE* **6**, e14745 (2011).
150. Gifford, R. & Tristem, M. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* **26**, 291–315 (2003).
151. Andersson, M.-L., Sjottem, E., Svineng, G. & Johansen, T. Comparative analyses of the LTRs of the ERV-H family of primatespecific, retrovirus-like elements isolated from marmoset, African green monkey, and man. *Virology* **234**, 14–30 (1997).
152. Shih, A., Coutavas, E. E. & Rush, M. G. Evolutionary implications of primate endogenous retroviruses. *Virology* **182**, 495–502 (1991).
153. Turner, G. *et al.* Insertional polymorphisms of full-length endogenous retroviruses in humans. *Curr. Biol.* **11**, 1531–1535 (2001).
154. Stoye, J. P. Endogenous retroviruses: still active after all this time? *Curr. Biol.* **11**, R914–R916 (2001).
155. Goff, S. P. Host factors exploited by retroviruses. *Nature Rev. Microbiol.* **5**, 253–263 (2007).
156. Hayward, W. S., Neel, B. G. & Astrin, S. M. Activation of a cellular *onc* gene by promoter insertion in ALV-induced lymphoid leukosis. *Nature* **290**, 475–480 (1981).
157. Jenkins, N. A., Copeland, N. G., Taylor, B. A. & Lee, B. K. Dilute (d) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. *Nature* **293**, 370–374 (1981).
158. Magiorkinis, G., Gifford, R. J., Katzourakis, A., De Ranter, J. & Belshaw, R. *Env*-less endogenous retroviruses are genomic superspreaders. *Proc. Natl Acad. Sci. USA* 23 Apr 2012 (doi:10.1073/pnas.1200913109).

Acknowledgements
I thank numerous colleagues in the retrovirology community and at the National Institute for Medical Research, London, UK, for many helpful discussions, and I apologize to those whose publications could not be cited here on account of space restrictions. Work in my laboratory is supported by the UK Medical Research Council (file reference U117512710).

Competing interests statement
The author declares no competing financial interests.

FURTHER INFORMATION
Jonathan P. Stoye's homepage:
http://www.nimr.mrc.ac.uk/research/jonathan-stoye

ALL LINKS ARE ACTIVE IN THE ONLINE PDF