Trends in Genetics

CellPress OPEN ACCESS

Review

Wildlife endogenous retroviruses: colonization, consequences, and cooption

Patric Jern [1,*] and Alex D. Greenwood [2,3,*]

Endogenous retroviruses (ERVs) are inherited genomic remains of past germline retroviral infections. Research on human ERVs has focused on medical implications of their dysregulation on various diseases. However, recent studies incorporating wildlife are yielding remarkable perspectives on long-term retrovirus–host interactions. These initial forays into broader taxonomic analysis, including sequencing of multiple individuals per species, show the incredible plasticity and variation of ERVs within and among wildlife species. This demonstrates that stochastic processes govern much of the vertebrate genome. In this review, we elaborate on discoveries pertaining to wildlife ERV origins and evolution, genome colonization, and consequences for host biology.

Highlights

A recent increase in the number of high-quality genome sequences and assemblies facilitates systematic screening of vertebrate DNA, better reflecting endogenous retrovirus (ERV) variability across wildlife and providing a means for evaluating historic and contemporary retrovirus transmission.

In contrast to commonly studied laboratory and domestic animals, recent studies of wildlife ERVs and exogenous retroviruses (XRVs) offer insights into historic and contemporary retrovirus activities in hosts undergoing natural selection.

Wildlife host populations under natural selection provide insights into ERV variation that segregates at various frequencies in host populations and highlight limitations of ERV detection relying on reference genome assemblies derived from one or a few individuals alone.

Phylogenetic differences between ERVs and contemporary XRVs suggest that XRVs are prone to extinction over evolutionary timescales, while the ERV record persists, or that many retroviruses remain to be discovered.

Large-scale phylogenomic screening for coopted ERV functions suggest that the frequency of coopted genes is underestimated.

Endogenous retroviruses across wildlife

Retrovirus and endogenous retrovirus association

Retroviral replication is unique among viruses, requiring the conversion of the RNA genome to DNA, which is integrated irreversibly as a provirus into the host genome. If the germline is infected, the resulting proviruses can be transmitted to offspring as heritable ERVs, which can be identified in descendant hosts long after the ancestor of the exogenous retrovirus (XRV) that infected the germline has become extinct (Box 1) [1]. Most ERV research has been in humans and laboratory mice [2–6]. However, wildlife (i.e., nondomesticated, either free-living or captive, species) share a long history with retrovirus germline infections, which provide a remarkable record of past virus–host interactions not observable in humans or mice. Since the initial release of the human genome assembly, the catalog of sequenced and annotated genomes from diverse species has grown rapidly with the development of current state-of-the-art sequencing technologies and increasingly sophisticated computational analyses [7–9] (see Figure 1).

Systematic screening of vertebrate DNA shows extreme variability in the ERV catalog across wildlife [10] and offers a glimpse into retrovirus evolution, patterns of host use, and mechanisms underlying retroviral transmission. Large-scale phylogenetic analyses utilizing reference XRVs and ERVs demonstrate an evolutionary spectrum, which can be divided into seven main genera (*alpha*-, *beta*-, *gamma*-, *delta*-, *epsilon*-, *lenti*-, and *spumavirus*) (Box 1). ERVs representing all main XRV genera have been detected and taxon-specific association patterns emerge [10–16]. Most ERVs are phylogenetically distant from reference XRVs, which suggests that retrovirus lineages are prone to extinction over evolutionary timescales, while the ERV record persists, or that many retroviruses remain to be discovered [10]. Phylogenomic studies show that ERVs from gamma- and betaretroviruses are the most abundant groups and, for example, primates have accumulated mostly gammaretroviral ERVs, whereas rodents have accumulated mostly betaretroviral ERVs, while simultaneously being prolific gammaretroviral spreaders [10,17]. Although phylogenetically basal ERVs can be associated with specific host species, such as old spuma-like ERVs in sloth and ancient gammaretroviral ERVs in insectivores, such as some bats, the origin of most ERV lineages remains unknown [14,18,19]. While ancient ERVs have

[1]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden
[2]Department of Wildlife Diseases, Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany
[3]School of Veterinary Medicine, Freie Unversität Berlin, Berlin, Germany

*Correspondence:
Patric.Jern@imbim.uu.se (P. Jern) and
greenwood@izw-berlin.de
(A.D. Greenwood).

Box 1. Endogenous retroviruses

Retroviruses are enveloped single-stranded RNA viruses that convert their genetic code into proviral DNA and integrate irreversibly into the host genome as part of their replication cycle. Although single stranded, each virion contains two independent copies of the viral genome. The provirus structure (~7–11 kb) is characterized by long terminal repeats (LTRs), which are identical at the time of conversion from RNA to DNA by the retroviral reverse transcriptase, flanking the main proviral genes encoding structural proteins (Gag), enzymes (Pro and Pol), and viral envelope proteins (Env) (Figure I). Proviral expression utilizing cellular RNA Polymerase II initiates in the upstream LTR and terminates in the downstream LTR. The canonical gene organization evolved to express required ratios of structural polyproteins and enzymes, with the Env polyproteins expressed through alternative RNA splicing. Additional small proteins and RNAs that facilitate replication of complex retroviruses are also generated by alternative splicing. Release of the assembled retrovirus from the cell is followed by extracellular maturation to infectious virus by proteolytic processing of polyproteins.

Retroviral germline infection may result in transmission across generations as an ERV that can be identified at varying frequencies in host populations, from single heterozygous copies in a single chromosome of an individual to homozygous integrations fixed in or among species. Although the ancestral infectious retrovirus may become extinct, the ERV can still be detected in descendant hosts, even after speciation. ERV integrations can have diverse effects on host genome function and structural ERV variations can impact host transcriptomes by providing novel promoter or enhancer sequences, alternative splicing from strong proviral splice donor and splice acceptor motifs, and termination of host gene transcription from proviral polyadenylation signals in the LTR sequences. Recombination between similar proviral or endogenous proviral segments can result in host genomic rearrangement and even loss of chromosomal DNA. More frequently, recombination occurs between the proviral LTRs, resulting in the loss of most of the proviral sequence and leaving a solo LTR at the site with retained promoter functions, such as transcription initiation or termination, or binding sites for cellular proteins involved in transcription or recruiting epigenetic signals for transcription regulation. ERV gene functions can be coopted by the host genome, altering host phenotypes, or ERVs can protect against infection by related exogenous retroviruses. The process can also be reversible because ERVs may complement defective proviruses *in trans*, or even recombine and produce novel exogenous retroviruses that can escape the germline.

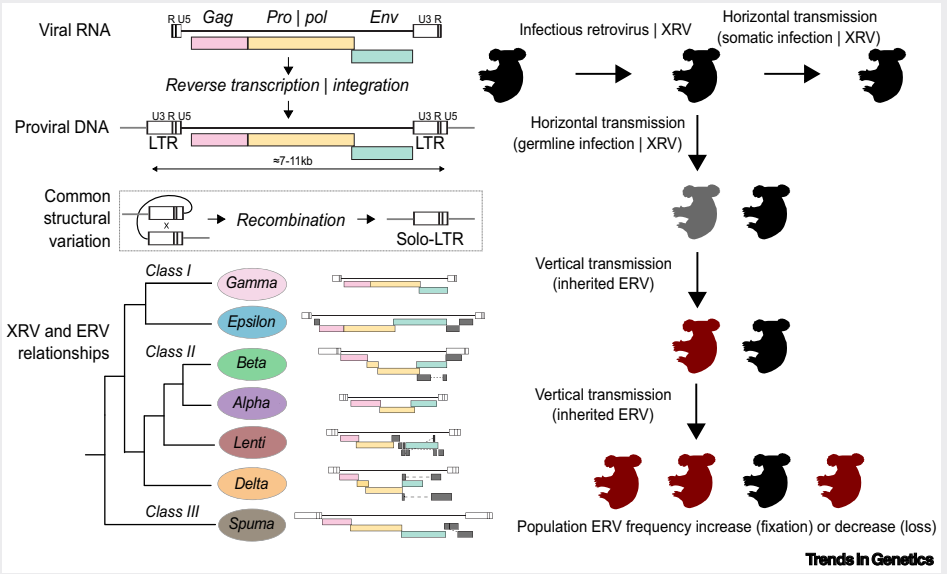

Figure I. Structure, classification, and transmission of endogenous and exogenous retroviruses. Abbreviations: ERV, endogenous retrovirus; LTR, long terminal repeat; XRV, exogenous retrovirus

been found in bats as a group, some bat lineages, such as the common vampire bat (*Desmodus rotundus*), appear to be exceptionally resistant to retroviral colonization and have low genomic ERV counts [20]. Therefore, accumulation and restriction have counteracting roles in ERV abundance and distribution.

Extended wildlife sampling facilitates evaluation of retrovirus transmission history by means of providing outgroup sequences for phylogenetic trees or demonstrating ancient cross-species

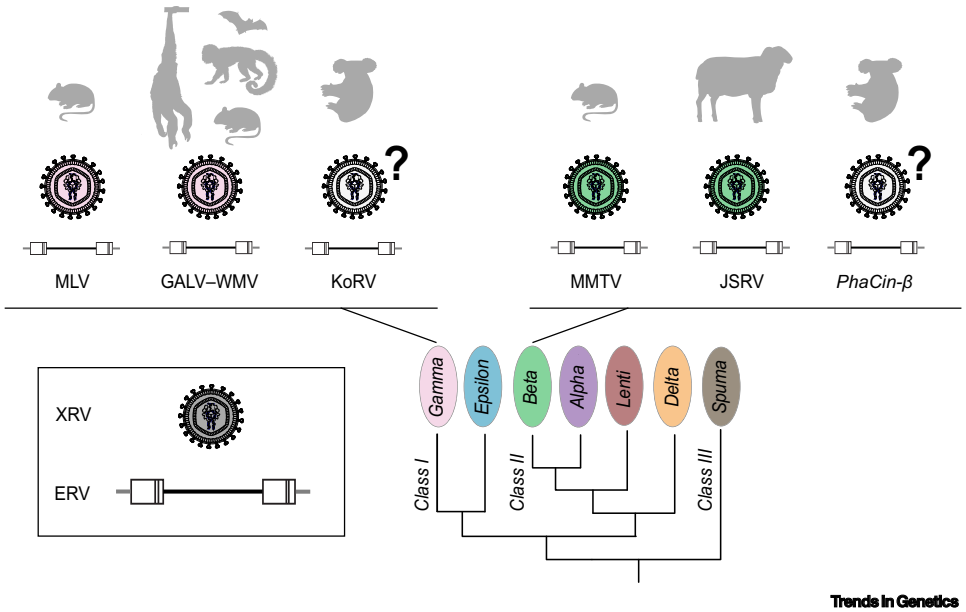

Figure 1. Endogenous retrovirus (ERV) germline colonization in diverse wildlife species. Retroviral germline colonization and known coexistence with contemporary exogenous retroviruses (XRVs) are dominated by the gamma- and betaretroviral lineages, which appear to be active in both mice [murine leukemia virus (MLV) and mouse mammary tumor virus (MMTV)] and koala [koala retrovirus (KoRV) and *PhaCin-β*]. Notably, multiple hosts appear to be permissive for infection and genome colonization by KoRV-related gibbon ape leukemia virus (GALV)-woolly monkey virus (WMV) sequences which appears to be relatively frequent. The '?' indicates that transmission of KoRV and *PhaCin-β* as XRV has not yet been experimentally established *in vivo,* unlike clear evidence for the other viral groups.

transmission when ERVs are included [10,14]. Wildlife host populations under natural selection provide insights into ERV variation that segregates at various frequencies in host populations and highlight limitations of ERV detection relying on reference genome assemblies derived from one or a few individuals alone [21,22]. While it is possible to estimate minimum ERV ages from detection along wildlife phylogeny, host lineage-specific ERVs and cross-species transmission present a challenge. Given the identical 5′ and 3′ long terminal repeat (LTR) sequences at the time of provirus integration (Box 1), LTR divergence and genomic mutation rate can provide age approximations, which improve with further sampling and analysis of ERV segregation at varying frequencies within host populations [21]. Wildlife ERVs also contribute to the repertoire of genomic markers used in genome-wide studies to identify disease associations [7]. Given that ERVs represent derived allelic variants upon integration (Box 1), they are distinct from commonly analyzed *de novo* single nucleotide polymorphisms (SNPs). Whereas ancestral states of SNPs are sometimes impossible to determine, ERV ancestry can be traced through wildlife history despite complete fixation in one or more host populations, which is not possible for completely fixed SNPs, regardless of sampling. Thus, ERV variation in wild populations can uncover evolutionary trajectories hidden from scrutiny in laboratory or domesticated species [21,23–26], including detection of cross-species transmission history.

Wildlife cross-species transmission

Examples of evolutionarily recent cross-species transmission are the koala (*Phascolarctos cinereus*) retrovirus (KoRV) and closely related gibbon ape leukemia virus (GALV). They are recent enough that very closely related XRVs are currently in circulation among wildlife. KoRV and GALV represent a puzzling association, given there has been extremely little faunal exchange over the

Wallace Line, a geographical barrier that separates most of South-East Asia from the Australo-Papua region [27]. Until recently, it was unclear how gibbons and koalas from either side of this barrier could share such closely related viruses. GALV is not found naturally in gibbons outside of the SEATO facility in Thailand [28], where it likely represents a laboratory infection of captive gibbons that spread to woolly monkeys within the same facility, yielding the five currently known GALV strains [GALV-X, GALV-Brain, GALV-SEATO, GALV-Hall's Island, and the woolly monkey virus (WMV)]. However, endogenous GALVs, specifically WMV related, were identified in melomid rodents in Indonesia and Australia, while closely related exogenous GALVs have been detected among bats from Papua New Guinea and Australia [29–32]. Two GALV-like lineages were identified in bats from China [*Hipposideros larvatus gammaretrovirus* (HlGRV) and *Rhinolophus hipposideros gammaretrovirus* (RhGRV)], which suggests that natural transmission across the Wallace Line is possible [30]. However, neither virus was completely sequenced; thus, it is unclear whether they are infectious or whether they are ERVs. However, none of the bats studied had GALV-KoRV ERVs, whereas the melomid WMVs are clearly endogenous [29]. While it appears that the ancestral GALV lineage is circulating in the Australo-Papua region, it remains unclear why endogenous GALV or exogenous KoRVs were not identified, especially given the recent colonization of koala genomes by KoRV [33]. However, it is possible that the exogenous ancestor retrovirus lineage is extinct. Other than undersampling of taxa in the region, it remains an open question why no endogenous KoRV-like sequences are found in other regional wildlife, despite extensive investigation, or why only WMV-like sequences or more divergent GALVs are found and none of the other GALV lineages [29,30,32].

Post-endogenization determinants of ERV germline colonization activity

Comparing different wildlife species also reveals patterns in viral evolution post germline colonization. A key determinant of ERV accumulation and genomic abundance within or among species is loss of *env* and *gag* coding potential exemplified by, for instance, intracisternal A-type retrotransposable elements (IAPs). Most widespread ERVs have accumulated mutations or deletions incompatible with the expression of functional Env proteins, suggesting possible transition to either an intracellular mode of replication or replication by complementation *in trans* from compatible *env*-competent proviruses elsewhere in the genome. For example, most retroviruses evolved an infection mode of replication or have lost functions as they became ERVs, and could then increase in copy numbers by intracellular replication [34]. In other cases, such as Env-encoding IAP (IAPE) elements, a specific mutation in the Gag N-terminal region causes resultant viral particles to bud in the endoplasmic reticulum instead of the cell membrane [35]. In both cases, cell-restricted elements appear to increase copy number to a higher extent compared with elements with intact genomes. Cell restriction and intracellular replication are likely key evolutionary turning points that determine ERV fitness and, therefore, few viruses exist as XRV and ERV simultaneously over long evolutionary periods.

Disruption of ERVs, particularly during the early stage of germline colonization, may be complex, affecting multiple viral genes. Recombinant KoRV variants (recKoRVs) have variable recombination breakpoints. In some cases, they represent a large proportion of the KoRVs in a given genome, the distribution of which varies among koalas geographically [36–38]. In contrast to disrupted *gag* or *env*, recKoRVs frequently exchange all coding regions with noncoding sequences. Structurally, they represent recombination events between KoRV and an ancient high-copy marsupial ERV called *Phascolarctos endogenous retrovirus* (PhER). Most recombinants identified (18 to date) result from microhomologies between KoRV and PhER [38]. Some recombinants, such as recKoRV1, have arisen multiple times independently in different koala populations and can represent 10% of the total KoRVs in a given koala. Surprisingly, some koala populations in South Australia, which lack KoRV, contain recKoRVs, and even recKoRV1-positive but KoRV-negative koalas have been identified [38,39].

Several aspects of retroviral transmission and germline colonization remain unresolved. It is not clear why some viral taxa are well represented as ERVs among vertebrates, whereas many other taxa, such as lentiviruses, appear only sporadically in wildlife genomes [10,15,40,41]. In some cases, the answer depends on receptor usage, such as lentivirus specificity to CD4 T cells and an inability to effectively utilize germ cell receptors, thus limiting opportunity for germline colonization [42]. In other cases, the barrier remains obscure, such as the relative absence of Jaagsiekte sheep retrovirus (JSRV)-like sequences in nonbovid hosts even though the receptor appears to be conserved [43].

Another unresolved question is what limits the copy number of ERVs, unlike other transposable elements. ERVs rarely exceed copy number in the low hundreds and most have <100 copies [44]. It is unclear what caps this number compared with intracellularly replicating retrotransposons, such as long or short interspersed nuclear elements (LINEs and SINEs, respectively), which may reach hundreds of thousands of copies per genome. In addition, comparing intracellularly replicating IAP with other IAP (see above), the copy number increase is orders of magnitude lower than for LINE and SINE retrotransposons, reflecting differences in adaptation to germline transmission.

Recent ERV colonizations and their consequences for host biology

Germline colonization

Not all vertebrates are equally likely to be infected by XRVs or colonized by ERVs. Body size has emerged as a potential contributing factor. Small-bodied vertebrates with early reproduction and multiple offspring per reproduction cycle are thought to tolerate a higher ERV burden and horizontal transmission of XRVs compared with large-bodied vertebrates [45]. It is possible that there are detrimental ERV-associated phenotypic manifestations; for example, neoplasia at any time before or during the reproductive life stage of larger bodied vertebrates will lower fitness and will be selected against. This is born out theoretically, with body size explaining 37% of the variance in ERV abundance, with large mammals having fewer integrations compared with smaller mammals. This also suggests that the mechanisms underlying colonization by ERVs differ; for example, IAPE success may be related to competition among IAPs, whereas colonization of large vertebrate genomes may be more influenced by selection pressure deriving from the host.

Most discovered ERVs are ancient and colonized their respective host germline millions of years ago, which has made it difficult to understand the underlying processes, such as effects on the host during colonization and changes in the viral population. However, there are a few small and large mammal examples of this process happening in real time. Among small mammals, examples generally derive from laboratory rodents of the genus *Mus* murine leukemia virus (MLV) and mouse mammary tumor virus (MMTV)-like viruses [46–48].

Among wild and domestic animals, JSRV is an example of a currently circulating XRV with ERV counterparts widely distributed among bovid hosts. Thought to be bovid specific, relatives of JSRV have been characterized in mouse lemurs from Madagascar, making the origin of the viral group unclear [49]. While some endogenous JSRVs (enJSRV) are old, several have been established over the past several hundred years and, one case, enJSRV-26, may have been established over the past 100–200 years [50].

Koalas currently represent the most tractable large vertebrate model for the process of ERV establishment in a host germline. The koala genome is currently being colonized by three retroviruses, the gammaretroviral KoRV and the two betaretroviral *PhaCin-β* and *PhaCin-β*-like lineages [26,51]. KoRV is the youngest and began colonizing the koala genome ~50 000 years ago. *PhaCin-β* is older, perhaps in the 100 000-years-ago range, whereas *PhaCin-β*-like represents

a young lineage relative to most known ERVs, but much older than both KoRV and *PhaCin-β*. For KoRV and *PhaCin-β*, koala populations are insertionally polymorphic, with no ERV loci fixed in frequency among koalas for KoRV and fewer than ten for *PhaCin-β*, based on current analyses. KoRV is not detected in some southern koalas, particularly on Kangaroo Island off the coast of South Australia. However, KoRV-negative koalas appear to have several copies of the recently described recKoRV [39]. However, the origin and health significance of these recKoRVs remain unclear. Understanding the distribution of KoRV integration sites and the sequences of widely shared integrations is instrumental for determining the adaptive process of KoRV to the koala host, as will similar future work be in terms of *PhaCin-β*.

Consequences of germline colonization for the host

There are clear health differences associated with KoRV distributions in koalas, with high KoRV loads in northern koalas being associated with higher rates of both neoplasia and *Chlamydia* compared with southern koalas, which lack intact KoRVs [52,53]. The link between prevalence and health is partly explained by KoRV variant expression and diversity, with recent minor variants, particularly in the *env* gene, emerging in large numbers, including novel exogenous KoRVs, such as KoRV-B [54–59]. How KoRV diversity and expression relate to disease progression is undetermined but demonstrates that, even as ERVs, variation in frequencies within a population can increase over short (50 000 years) evolutionary timescales.

The negative health effects of germline integrations by retroviruses may place ERVs under evolutionary pressure to become attenuated, particularly in large mammals. While debated associations with disease are common for human ERVs, with prominent examples in cancer, multiple sclerosis, and various other diseases [2], it was unclear until recently whether these are consequences of changes that happened long after ERV establishment in germline or are unresolved virus host co-adaptions. KoRV is instructive in this regard. Koalas have a tenfold higher rate of leukemia and lymphoma compared with other mammals, which is a typical manifestation of *gammaretrovirus* infection related to MLV [60]. However, the association between KoRV and neoplasia was merely correlative, until McEwen *et al.* recently demonstrated that KoRV-A can lead to neoplasia by three mechanisms common to exogenous retroviruses [61]. First, accumulation of KoRV integrations in somatic cells occurs preferentially in known oncogenes. Second, germline integrations of KoRV in oncogenes can lead to a predisposition to specific heritable cancers. Third, transduction of the full-length BCL2XL oncogene was observed and is reminiscent of one of the earliest described retroviruses, Rous sarcoma virus (RSV), which was found to transduce the *src* gene [62]. In all cases, the result of KoRV integration or transduction is often increased expression of the disrupted oncogene. Therefore, at the earliest stages of ERV colonization, the manifestations of young ERVs are similar to those expected for related XRVs, with the additional burden of every cell in the host being a potential source of infection or mutations derived from retroviral activity. This could partly explain the evolution of host countermeasures against XRVs and ERVs, and rapid viral degradation over time as seen, for example, with recKoRVs (see following text).

A correlative link exists between KoRV and *Chlamydia* in koalas, which predominates in northern high KoRV-burdened populations and is less common or problematic in southern koalas with lower or even no KoRV burden [52]. This is an intriguing correlation given that KoRV, similar to most gammaretroviruses, has an immunosuppressive domain (p15E) in the Env protein that could explain the inability of many populations in Queensland and New South Wales to control chlamydial infections [63]. However, the causal or mechanistic link between *Chlamydia* and KoRV remains undetermined and the involvement of integrations, expression of KoRV, or mechanisms unrelated to KoRV remains speculative.

Host responses

Given their long-term genomic associations, restriction of transposable elements has evolved over millions of years and can be species specific. Piwi-interacting RNAs (piRNAs) are anti-sense RNAs that normally inhibit transposon expression *in trans* by generating anti-sense homologs of the transposons they suppress and sequestering the expressed RNA, thereby preventing transposition [64]. However, during KoRV infection, sense strand piRNAs are generated from unspliced KoRV transcripts, which is thought to halt replication *in cis* until anti-sense piRNAs are generated that can then suppress KoRV expression *in trans* [64]. However, suppression is leaky because full-length transcripts of KoRV genomes and proteins are abundantly detected in koala tissue despite piRNA inhibition. These observations likely reflect the recent genomic colonization by KoRV and the lack of a fully functional piRNA-silencing system.

Several enJSRVs have recently colonized bovid germlines and are restricted to the *Ovis* lineage, to which domestic sheep, *Ovis aries*, belong. Two enJSRV (enJS56A1 and envJSRV-20) were found to be restricted in *O. aries*, but not in other *Ovis* spp. A tryptophan mutation at the 21st position (W21) of the Gag protein was observed, whereas other *Ovis* generally had an arginine at the same position or, in many cases, even lacked the ERVs [50]. The W21 mutation is transdominant and prevents replication of similar JSRV-like retroviruses elsewhere in the genome, thus benefiting the host. However, an ERV identified in one sheep specimen (envJSRV-26) that formed within the past 200 years was able to escape the W21 restriction. Thus, JSRVs remain variable and colonize genomes of sheep and their relatives, although mutations within the ERVs may protect the host from further germline infection.

Germline escape

As recKoRVs demonstrate, ERVs and XRVs can interact and generate novel viral sequences. As in recKoRVs, these interactions could result in non-infectious retroviruses lacking coding potential for one or more viral proteins. However, infection by XRVs may mobilize ERVs or, in some cases, ERVs can recombine to generate novel XRVs. In the former case, in mice infected with ecotropic (i.e., only infecting murine cells) MuLV, the latter recombined with endogenous polytropic MLVs, generating novel polytropic XRVs [65]. In addition, these novel recombinants were able to mobilize full-length polytropic ERVs, demonstrating the ease with which ERV sequences can be transmitted as XRVs long after the ERV has established in the host. In another case, XMRV, a virus initially and erroneously identified as a zoonotic rodent retrovirus, was later demonstrated to have resulted from four recombination events between two murine ERVs (preXMRV-1 and preXMRV-2) in nude mouse cell cultures [66]. The novel XRV was able to infect human prostate cells in culture, leading to the erroneous conclusion that patients with prostate cancer had been infected with a zoonotic retrovirus. This finding demonstrates that an ERV is not necessarily an endpoint for the retrovirus lineage but represents a continuous process, which may also cloud the origins of retroviruses at deeper taxonomic levels if different parts of current viral genomes derive from multiple exchanges between viral groups over millions of years. For example, many betaretroviruses have incorporated gammaretroviral sequences, especially the *env* gene [51,67–69].

Cooption

ERVs have been repurposed as novel host genes, such as syncytins (see following text), to contribute to host defense against infection, and to provide platforms for modulating host gene expression and evolution (Figure 2). Research on ERVs as genomic regulators dates back to the mid-20th century and high-quality genome data and broad sampling across diverse hosts now enable in-depth associations among retroviruses, ERVs, and their hosts. Both the control regions and retroviral genes of ERVs can be coopted. Currently, evidence for cooption is restricted

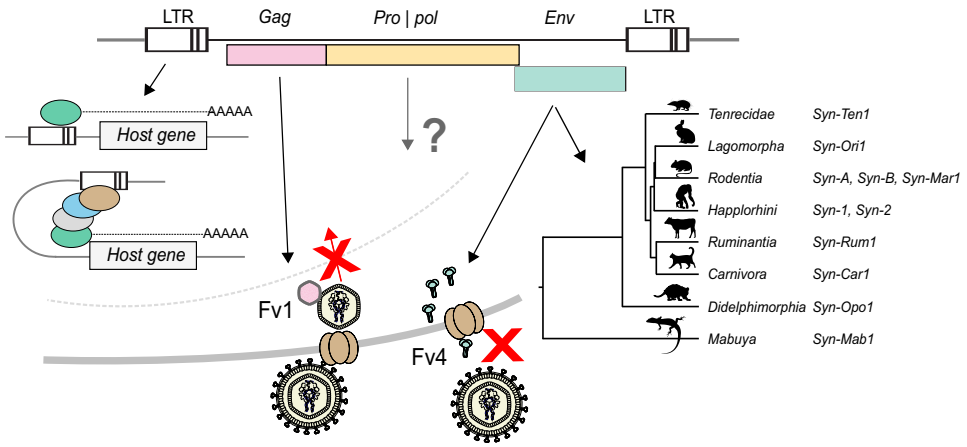

Figure 2. Endogenous retrovirus (ERV) cooption by wildlife. Individual proviral functional sequences are shown with respect to observed cooption mechanisms. Long terminal repeats (LTRs) influence adjacent host gene transcription by providing novel promoters, recruiting cellular proteins, or acting as transcription enhancers altering host gene expression. The ERV *gag*-derived *Fv1* interferes with uncoating of retroviruses, preventing viral DNA entry to the cell nucleus and, thus, restricting retroviral replication. The ERV *env*-derived *Fv4* interacts with cell surface receptors, blocking access to them by related exogenous retroviruses (XRVs) to prevent binding for virus–host membrane fusion and viral cell entry. *Env*-derived syncytins have been coopted independently from different proviruses across a broad host phylogeny, even in viviparous animals, suggesting development of a primitive placenta followed by multiple independent improvements enabled by retroviral infections and germline colonizations.

to the *gag* and *env* genes and LTRs. Cooption of *pol* and *pro* is possible but has not yet been described. LTR cooption is widespread and well known. LTRs near host genes may override, replace, or act as enhancer elements to host gene promoters, thereby putting host genes under retroviral expression control [70]. In some cases, LTRs with binding sites for major control proteins, such as P53, may integrate near genes that place entire networks of genes under master regulatory control in species-specific ways [71]. The ability to provide binding platforms for cellular proteins contributes to rewired host gene transcription control, such as interacting with TRIM28 proteins that recruit epigenetic signals for ERV silencing, which, when disrupted, reactivates ERVs and consequently modulates their effects on adjacent genes involved in, for instance, neural development [72–74].

Cooption of *gag* genes has been associated with host defense against XRVs among rodents (*Fv1*) (Figure 2). The *Fv1* locus prevents production of functional capsid proteins by exogenous MLVs [75]. Similarly, the *Fv4* locus represents a coopted MLV *env* gene, the product of which localizes to the cellular receptor used by exogenous MLVs and blocks cellular entry of related MLVs [76,77]. Cooption of *gag* genes is also associated with intracellular RNA transfer in the nervous system (ARC genes) [78].

While the cooption examples above are not strictly assimilation of retroviral function, syncytins are evidence of retroviral gene function domesticated by the host. Syncytins are a variety of largely species-specific occurrences of ERV Env proteins that were acquired for their fusogenic properties during placental development. The contributing ERVs are not shared among mammals, suggesting that retroviral domestication occurred independently by convergent evolution on multiple occasions [79]. It even extends to include viviparous reptiles, such as *Mabuya* lizards, which also develop a placenta (Figure 2) [80]. One possible contributing factor to why few ERVs have been coopted for their fusogenic function, whereas tens of thousands of other

ERVs are not, is that these ERVs may be primed for regulated expression. Enrichment of binding motifs for cellular proteins, such as the domesticated mammal-specific zinc-finger BED-type domain containing 6 (ZEBD6) DNA transposon, near syncytin genes across species, may be one such priming factor [81]. Although a direct connection has not been determined, it is possible that enriched transposon control facilitates cooption of specific ERVs. Variable expression and the cooption of different retroviral Env proteins may explain differences in placental structures among vertebrates [82].

Recent results of large-scale phylogenomic screens for coopted ERV function suggest that the existing known repertoire of coopted genes, *gag* and *env* specifically, underestimate the frequency of cooption [83]. A total of 177 putative gene cooptions, 93 *gag* and 84 *env*, were identified across a broad taxonomic sample. While these promising results await functional confirmation and characterization, the increasing number of wildlife genomes will certainly expand these numbers considerably.

Concluding remarks

Wildlife are the key to understanding the origin, diversity, and distribution of ERVs (see Outstanding questions). Studies of wildlife lineages reveal how different species have dealt with extant viruses at the time of germline colonization. They are our only resource for understanding the natural germline colonization process and its health consequences. Finally, wildlife allows us to identify and characterize examples of cooption of retroviral function. Given that only a relatively small number of wildlife genomes have been sequenced but that more are being continuously added to the databases, more detailed accounts of processes identified are expected to emerge, as will completely unexpected phenomena.

Outstanding questions

What are the general evolutionary patterns of host adaptation to ERVs and the detailed mechanisms underlying retroviral transmission?

What are the source of ERV lineages and the mechanism of germline colonization?

What caps ERV copy numbers in host DNA to a few or ~100 copies per ERV lineage, unlike intracellularly replicating retrotransposons, which sometime occur in thousands of copies?

Why are some viral taxa well represented as ERVs among vertebrates but many other taxa appear only sporadically in wildlife genomes?

What determines co-transmission of co-occurring pathogens, such as *Chlamydia*, associated with retrovirus infection?

Acknowledgments

P.J. was supported by the Swedish Research Council (grant 2021-01740). A.D.G. was supported by grants GR 3924/15-1 and GR 3924/12-1 from the Deutsche Forschungsgemeinschaft (DFG). The authors thank Rachael Tarlinton for commenting on an early draft of the manuscript.

Declaration of interests

No interests are declared.

References

1. Johnson, W.E. (2019) Origins and evolutionary consequences of ancient endogenous retroviruses. *Nat. Rev. Microbiol.* 17, 355–370
2. Jansz, N. and Faulkner, G.J. (2021) Endogenous retroviruses in the origins and treatment of cancer. *Genome Biol.* 22, 147
3. Kozak, C.A. (2015) Origins of the endogenous and infectious laboratory mouse gammaretroviruses. *Viruses Basel* 7, 1–26
4. Latifi, T. *et al.* (2022) The role of human endogenous retroviruses (HERVs) in multiple sclerosis and the plausible interplay between HERVs, Epstein-Barr virus infection, and vitamin D. *Mult. Scler. Relat. Dis.* 57, 103318
5. Pisano, M.P. *et al.* (2019) Comprehensive characterization of the human endogenous retrovirus HERV-K(HML-6) group: overview of structure, phylogeny, and contribution to the human genome. *J. Virol.* 93, e00110–19
6. Xue, B. *et al.* (2020) Human endogenous retrovirus K (HML-2) in health and disease. *Front. Microbiol.* 11, 1690
7. Christmas, M.J. *et al.* (2023) Evolutionary constraint and innovation across hundreds of placental mammals. *Science* 380, eabn3943
8. Giani, A.M. *et al.* (2020) Long walk to genomics: history and current approaches to genome sequencing and assembly. *Comput. Struct. Biotec* 18, 9–19
9. Osmanski, A.B. *et al.* (2023) Insights into mammalian TE diversity through the curation of 248 genome assemblies. *Science* 380, eabn1430
10. Hayward, A. *et al.* (2015) Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. *Proc. Natl. Acad. Sci. U. S. A.* 112, 464–469
11. Aiewsakun, P. *et al.* (2019) Modular nature of simian foamy virus genomes and their evolutionary history. *Virus Evol.* 5, vez032
12. Chu, L.Y. *et al.* (2023) Jawless vertebrates do not escape retrovirus infection. *Virology* 583, 52–55
13. Hron, T. *et al.* (2019) Deltaretroviruses have circulated since at least the Paleogene and infected a broad range of mammalian species. *Retrovirology* 16, 33
14. Katzourakis, A. *et al.* (2009) Macroevolution of complex retroviruses. *Science* 325, 1512
15. Katzourakis, A. *et al.* (2007) Discovery and analysis of the first endogenous lentivirus. *Proc. Natl. Acad. Sci. U. S. A.* 104, 6261–6265
16. Wang, X.J. *et al.* (2023) Identification of cartilaginous fish endogenous foamy virus rooting to vertebrate counterparts. *J. Virol.* 97, e0181622
17. Hayward, A. *et al.* (2013) Broad-scale phylogenomics provides insights into retrovirus-host evolution. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20146–20151
18. Cui, J. *et al.* (2015) Bats and rodents shape mammalian retroviral phylogeny. *Sci. Rep.* 5, 16561
19. Slater, G.J. *et al.* (2016) Evolutionary relationships among extinct and extant sloths: the evidence of mitogenomes and retroviruses. *Genome Biol. Evol.* 8, 607–621

20. Mendoza, M.L.Z. *et al.* (2018) Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. *Nat. Ecol. Evol.* 2, 659–668
21. Hill, J. *et al.* (2022) Spatiotemporal variations in retrovirus-host interactions among Darwin's finches. *Nat. Commun.* 13, 6033
22. Rivas-Carrillo, S.D. *et al.* (2018) Whole-genome comparison of endogenous retrovirus segregation across wild and domestic host species populations. *Proc. Natl. Acad. Sci. U. S. A.* 115, E12465
23. Halo, J.V. *et al.* (2019) Origin and recent expansion of an endogenous gammaretroviral lineage in domestic and wild canids. *Retrovirology* 16, 6
24. Holloway, J.R. *et al.* (2019) Gorillas have been infected with the HERV-K (HML-2) endogenous retrovirus much more recently than humans and chimpanzees. *Proc. Natl. Acad. Sci. U. S. A.* 116, 1337–1346
25. Pettersson, M.E. and Jern, P. (2019) Whole-genome analysis of domestic chicken selection lines suggests segregating variation in ERV makeups. *Genes-Basel* 10, 162
26. Tarlinton, R.E. *et al.* (2006) Retroviral invasion of the koala genome. *Nature* 442, 79–81
27. Ali, J.R. and Heaney, L.R. (2021) Wallace's line, Wallacea, and associated divides and areas: history of a tortuous tangle of ideas and labels. *Biol. Rev.* 96, 922–942
28. Siegal-Willott, J.L. *et al.* (2015) Evaluation of captive gibbons (*Hylobates* spp., *Nomascus* spp., *Symphalangus* spp.) in North American zoological institutions for gibbon ape leukemia virus (Galv). *J. Zoo Wildlife Med.* 46, 27–33
29. Alfano, N. *et al.* (2016) Endogenous gibbon ape leukemia virus identified in a rodent (*Melomys burtoni* subsp.) from Wallacea (Indonesia). *J. Virol.* 90, 8169–8180
30. Hayward, J.A. *et al.* (2020) Infectious KoRV-related retroviruses circulating in Australian bats. *Proc. Natl. Acad. Sci. U. S. A.* 117, 9529–9536
31. McMichael, L. *et al.* (2019) A novel Australian flying-fox retrovirus shares an evolutionary ancestor with koala, gibbon and *Melomys* gamma-retroviruses. *Virus Genes* 55, 421–424
32. Simmons, G. *et al.* (2014) Discovery of a novel retrovirus sequence in an Australian native rodent (*Melomys burtoni*): a putative link between gibbon ape leukemia virus and koala retrovirus. *PLoS ONE* 9, e106954
33. Ishida, Y. *et al.* (2015) Proliferation of endogenous retroviruses in the early stages of a host germ line invasion. *Mol. Biol. Evol.* 32, 109–120
34. Magiorkinis, G. *et al.* (2012) Env-less endogenous retroviruses are genomic superspreaders. *Proc. Natl. Acad. Sci. U. S. A.* 109, 7385–7390
35. Ribet, D. *et al.* (2008) An infectious progenitor for the murine IAP retrotransposon: emergence of an intracellular genetic parasite from an ancient retrovirus. *Genome Res.* 18, 597–609
36. Hobbs, M. *et al.* (2017) Long-read genome sequence assembly provides insight into ongoing retroviral invasion of the koala germline. *Sci. Rep.* 7, 15838
37. Johnson, R.N. *et al.* (2018) Adaptation and conservation insights from the koala genome. *Nat. Genet.* 50, 1102-+
38. Lober, U. *et al.* (2018) Degradation and remobilization of endogenous retroviruses by recombination during the earliest stages of a germ-line invasion. *Proc. Natl. Acad. Sci. U. S. A.* 115, 8609–8614
39. Tarlinton, R.E. *et al.* (2022) Differential and defective transcription of koala retrovirus indicates the complexity of host and virus evolution. *J. Gen. Virol.* 103, 001749
40. Gifford, R.J. *et al.* (2008) A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proc. Natl. Acad. Sci. U. S. A.* 105, 20362–20367
41. Han, G.Z. and Worobey, M. (2015) A primitive endogenous lentivirus in a colugo: insights into the early evolution of lentiviruses. *Mol. Biol. Evol.* 32, 211–215
42. Clapham, P.R. and McKnight, A. (2001) HIV-1 receptors and cell tropism. *Br. Med. Bull.* 58, 43–59
43. Rai, S.K. *et al.* (2001) Candidate tumor suppressor HYAL2 is a glycosylphosphatidylinositol (GPI)-anchored cell-surface receptor for Jaagsiekte sheep retrovirus, the envelope protein of which mediates oncogenic transformation. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4443–4448
44. Belshaw, R. *et al.* (2005) High copy number in human endogenous retrovirus families is associated with copying mechanisms in addition to reinfection. *Mol. Biol. Evol.* 22, 814–817
45. Katzourakis, A. *et al.* (2014) Larger mammalian body size leads to lower retroviral activity. *PLoS Pathog.* 10, e1004214
46. Baillie, G.J. *et al.* (2004) Multiple groups of endogenous betaretroviruses in mice, rats, and other mammals. *J. Virol.* 78, 5784–5798
47. Nellaker, C. *et al.* (2012) The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol.* 13, R45
48. Young, G.R. *et al.* (2012) Resurrection of endogenous retroviruses in antibody-deficient mice. *Nature* 491, 774–778
49. Kessler, S.E. *et al.* (2023) Long-term host-pathogen evolution of endogenous beta- and gammaretroviruses in mouse lemurs with little evidence of recent retroviral introgression. *Virus Evol.* 9, veac117
50. Arnaud, F. *et al.* (2008) Coevolution of endogenous Betaretroviruses of sheep and their host. *Cell. Mol. Life Sci.* 65, 3422–3432
51. Lillie, M. *et al.* (2022) Expansion of a retrovirus lineage in the koala genome. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2201844119
52. Quigley, B.L. and Timms, P. (2020) Helping koalas battle disease - recent advances in *Chlamydia* and koala retrovirus (KoRV) disease understanding and treatment in koalas. *FEMS Microbiol. Rev.* 44, 583–605
53. Zheng, H.Q. *et al.* (2020) Koala retrovirus diversity, transmissibility, and disease associations. *Retrovirology* 17, 34
54. Blyton, M.D.J. *et al.* (2022) Koala retrovirus load and non-A subtypes are associated with secondary disease among wild northern koalas. *PLoS Pathog.* 18, e1010513
55. Blyton, M.D.J. *et al.* (2022) Geographic patterns of koala retrovirus genetic diversity, endogenization, and subtype distributions. *Proc. Natl. Acad. Sci. U. S. A.* 120, e2220541120
56. Chappell, K.J. *et al.* (2017) Phylogenetic diversity of koala retrovirus within a wild koala population. *J. Virol.* 91, e01820
57. Joyce, B.A. *et al.* (2022) Diversity and transmission of koala retrovirus: a case study in three captive koala populations. *Sci. Rep.* 12, 15787
58. Joyce, B.A. *et al.* (2021) Koala retrovirus genetic diversity and transmission dynamics within captive koala populations. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2024021118
59. Xu, W.Q. *et al.* (2013) An exogenous retrovirus isolated from koalas with malignant neoplasias in a US zoo. *Proc. Natl. Acad. Sci. U. S. A.* 110, 11547–11552
60. Fan, H. (1997) Leukemogenesis by Moloney murine leukemia virus: a multistep process. *Trends Microbiol.* 5, 74–82
61. McEwen, G.K. *et al.* (2021) Retroviral integrations contribute to elevated host cancer rates during germline invasion. *Nat. Commun.* 12, 1316
62. Swanstrom, R. *et al.* (1983) Transduction of a cellular oncogene - the genesis of Rous-sarcoma virus. *Proc. Natl. Acad. Sci. U. S. A.* 80, 2519–2523
63. Ishida, Y. *et al.* (2015) Sequence variation of koala retrovirus transmembrane protein p15E among koalas from different geographic regions. *Virology* 475, 28–36
64. Yu, T.X. *et al.* (2019) The piRNA response to retroviral invasion of the koala genome. *Cell* 179, 632–643
65. Evans, L.H. *et al.* (2009) Mobilization of endogenous retroviruses in mice after infection with an exogenous retrovirus. *J. Virol.* 83, 2429–2435
66. Paprotka, T. *et al.* (2011) Recombinant origin of the retrovirus XMRV. *Science* 333, 97–101
67. Chiu, C.N. *et al.* (2003) Exchange of genetic sequences of long terminal repeat and the env gene by a promiscuous primate type D retrovirus. *Virus Res.* 96, 107–111
68. Escalera-Zamudio, M. *et al.* (2015) A novel endogenous Betaretrovirus in the common vampire bat (*Desmodus rotundus*) suggests multiple independent infection and cross-species transmission events. *J. Virol.* 89, 5180–5184
69. Sonigo, P. *et al.* (1986) Nucleotide-sequence of Mason-Pfizer monkey virus - an immunosuppressive D-type retrovirus. *Cell* 45, 375–385
70. Durnaoglu, S. *et al.* (2021) Human endogenous retroviruses as gene expression regulators: insights from animal models into human diseases. *Mol. Cells* 44, 861–878

71. Wang, T. *et al.* (2007) Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc. Natl. Acad. Sci. U. S. A.* 104, 18613–18618
72. Enriquez-Gasca, R. *et al.* (2023) Co-option of endogenous retroviruses through genetic escape from TRIM28 repression. *Cell Rep.* 42, 112625
73. Fasching, L. *et al.* (2015) TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. *Cell Rep.* 10, 20–28
74. Jonsson, M.E. *et al.* (2021) Activation of endogenous retroviruses during brain development causes an inflammatory response. *EMBO J.* 40, e106423
75. Boso, G. *et al.* (2018) Ancient evolutionary origin and positive selection of the retroviral restriction factor Fv1 in muroid rodents. *J. Virol.* 92, e00850
76. Boso, G. *et al.* (2021) Patterns of coevolutionary adaptations across time and space in mouse gammaretroviruses and three restrictive host factors. *Viruses* 13, 1864
77. Malfavon-Borja, R. and Feschotte, C. (2015) Fighting fire with fire: endogenous retrovirus envelopes as restriction factors. *J. Virol.* 89, 4047–4050
78. Pastuzyn, E.D. *et al.* (2018) The neuronal gene Arc encodes a repurposed retrotransposon Gag protein that mediates intercellular RNA transfer. *Cell* 172, 275–288
79. Imakawa, K. *et al.* (2015) Baton pass hypothesis: successive incorporation of unconserved endogenous retroviral genes for placentation during mammalian evolution. *Genes Cells* 20, 771–788
80. Cornelis, G. *et al.* (2017) An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental *Mabuya* lizard. *Proc. Natl. Acad. Sci. U. S. A.* 114, E10991–E11000
81. Pettersson, M.E. and Jern, P. (2021) ZBED6 binding motifs correlate with endogenous retroviruses and Syncytin genes. *Virus Evol.* 7, veaa083
82. Funk, M. *et al.* (2019) Capture of a hyena-specific retroviral envelope gene with placental expression associated in evolution with the unique emergence among carnivorans of hemochorial placentation in Hyaenidae. *J. Virol.* 93, e01811-18
83. Wang, J.H. and Han, G.Z. (2020) Frequent retroviral gene co-option during the evolution of vertebrates. *Mol. Biol. Evol.* 37, 3232–3242