RESEARCH ARTICLE

Cross-Species Transmission and Differential Fate of an Endogenous Retrovirus in Three Mammal Lineages

Xiaoyu Zhuo, Cédric Feschotte*

Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah, United States of America

* cedric@genetics.utah.edu

Abstract

Endogenous retroviruses (ERVs) arise from retroviruses chromosomally integrated in the host germline. ERVs are common in vertebrate genomes and provide a valuable fossil record of past retroviral infections to investigate the biology and evolution of retroviruses over a deep time scale, including cross-species transmission events. Here we took advantage of a catalog of ERVs we recently produced for the bat *Myotis lucifugus* to seek evidence for infiltration of these retroviruses in other mammalian species (>100) currently represented in the genome sequence database. We provide multiple lines of evidence for the cross-ordinal transmission of a gammaretrovirus endogenized independently in the lineages of vespertilionid bats, felid cats and pangolin ~13–25 million years ago. Following its initial introduction, the ERV amplified extensively in parallel in both bat and cat lineages, generating hundreds of species-specific insertions throughout evolution. However, despite being derived from the same viral species, phylogenetic and selection analyses suggest that the ERV experienced different amplification dynamics in the two mammalian lineages. In the cat lineage, the ERV appears to have expanded primarily by retrotransposition of a single proviral progenitor that lost infectious capacity shortly after endogenization. In the bat lineage, the ERV followed a more complex path of germline invasion characterized by both retrotransposition and multiple infection events. The results also suggest that some of the bat ERVs have maintained infectious capacity for extended period of time and may be still infectious today. This study provides one of the most rigorously documented cases of cross-ordinal transmission of a mammalian retrovirus. It also illustrates how the same retrovirus species has transitioned multiple times from an infectious pathogen to a genomic parasite (i.e. retrotransposon), yet experiencing different invasion dynamics in different mammalian hosts.

Author Summary

The cross-species transmission of viruses poses a continuous threat to public health. Bats are increasingly recognized as a major reservoir for zoonotic RNA viruses, including

rabies, Ebola, and possibly MERS, but little is known about their capacity to harbor and transmit retroviruses. Here we investigated past incidents of cross-species transmission involving bat retroviruses, by screening for the presence of endogenous retroviruses (ERVs) previously identified in the genome of the little brown bat in more than 100 diverse mammal species. This screen revealed an intriguing case of a gammaretrovirus that independently infiltrated the germ line of species belonging to three mammalian orders: vesper bat, felid cat and pangolin. We found that the ERV initiated its genomic invasion of the three lineages around the same timeframe ~13–25 million years ago, but experienced a different fate in each lineage. In the pangolin lineage, the ERV's genomic propagation stalled shortly after endogenization, while it amplified continuously through-out felid and vesper bat evolution to generate hundreds of species-specific insertions in each lineage. Furthermore, in the cat lineage genomic amplification appears to have occurred predominantly via retrotransposition; while in bats the ERV has expanded via a mixture of retrotransposition and reinfection activity that may still be ongoing.

Introduction

Viral cross-species transmission (CST) represents a major threat to both human and animal populations. Most viral diseases of humans are zoonotic: they stem from CST of viruses from domestic or wild animals [1]. The explosion and development of human society, including modern transportation, over the last 100 years has exposed us to an increasing number of pathogens [2]. AIDS, which has caused more than 25 million deaths over the past ~30 years (aids. gov), is one of the most notorious examples of a pandemic initiated by viral CST [3,4]. The pathogens causing AIDS (HIV-1 and HIV-2) are retroviruses, a family of RNA viruses that use reverse transcription to replicate their genome [5]. Other retroviral CST events have been documented within primates, felids and ruminants, suggesting that retroviral CST represents a continuous threat to human and animal health [6–10].

Retroviruses are unique amongst animal viruses in that chromosomal integration of so-called proviruses is an obligatory step in their replication cycle [5]. As a consequence, retroviral infection of germ cells or their progenitors result in proviruses that may be vertically inherited along with the host genome. Such inheritable proviruses are called endogenous retroviruses (ERVs). Under some circumstances, which are still poorly understood, ERVs can further propagate within the genome and spread in the population, resulting in the formation of large families of interspersed repeats in the host genome [11]. Despite the potentially deleterious consequences associated with the genomic propagation of ERVs, the process has been remarkably pervasive during mammalian evolution. Indeed every mammalian genome thus far examined harbor a great abundance and diversity of ERVs, which are mostly lineage-specific. For example, 8% of the human genome is composed of ERV sequences derived from a wide variety of retroviruses acquired at different time points during primate evolution [12–14]. Once integrated and endogenized, most ERVs appear to evolve at the host's neutral mutation rate, which is much slower than the mutation rate of exogenous retroviruses (XRVs) [15]. Therefore ERVs provide a valuable fossil record of past retroviral infections and a unique opportunity to investigate retroviral evolution at a deep time scale, including CST events [16–20].

Many ancient CST events have been inferred by comparing ERV sequences across species [21–28]. Most of the well-documented cases of retroviral CST events involve closely related host species (e.g. from the same order). Indeed, it is thought that viral CST is often constrained by the evolutionary distance between donor and recipient species [19,20,29]. The observation

that all retroviruses known to infect humans have been acquired from other primates is consistent with this notion [7]. However, retroviral CST events can also occur between distantly related species. For example, the cat RD114 gammaretrovirus is a recombinant containing an envelope domain mostly closely related to Baboon endogenous virus (BaEV), and is thought have been acquired by the domestic cat from an Old World monkey [30,31]. Also, the koala retrovirus (KoRV), which is currently spreading and undergoing endogenization in the wild, is very closely related to gibbon ape leukemia virus (GALV) and to ERVs found in Asian rodents, from which it was most likely acquired [32]. It has also been reported that reticuloendotheliosis virus (REV) was likely transmitted from mammals to birds [10]. Recent phylogenomics surveys of ERVs across a wide variety of vertebrate species suggested that CST between widely diverged species (i.e. from different orders or classes) may be more common than initially anticipated [19,20,33,34]. However, the evidence remains limited and more detailed case studies are needed to confirm this idea.

Bats (order Chiroptera) are increasingly regarded as exceptionally potent reservoirs of zoonotic viruses [35–40]. Indeed, a variety of bat species have been implicated in the spillover of diverse and highly pathogenic RNA viruses such as Rabies, Nipah, Hendra, SARS, Marburg, and Ebola viruses in the human population [41]. Very recently, one potential case of CST of an endogenous betaretrovirus involving phyllostomid bats, rodents and New World monkeys was reported [28]. We previously produced a comprehensive catalog of ERVs in the vespertilionid bat *Myotis lucifugus* [42] (referred to as MLERVs hereafter), documenting a rich and recent history of retroviral infections in this species lineage. Here, we have taken advantage of this resource to seek evidence of CST events implicating MLERVs. We identified an intriguing case of a gammaretrovirus that colonized independently the genomes of vespertilionid bats, felids and pangolin but followed a different fate and amplification dynamics in these lineages.

Results

ERVs closely related to MLERV1 are present in species from three mammal orders

To detect possible CST events involving *M. lucifugus* ERVs, we used the sequence of the reverse transcriptase domain (RVT_1) (642 nt) from members of each of the 86 MLERV subfamilies previously identified [42] as queries in megaBLAST searches of all mammal genomes deposited in the NCBI whole genome shotgun (WGS) database as of February 2015 (107 mammal species). Excluding hits to *M. lucifugus*, the most significant hits (>80% nucleotide identity over the entire domain; e-value $< 10^{-80}$) were obtained with a query representing the MLERV1 family [42] against the genome assemblies of the domestic cat (*Felis catus*) [43], Amur tiger (*Panthera tigris*) [44] and Chinese pangolin (*Manis pentadactyla*). In addition, and less surprisingly, many highly significant hits to MLERV1 were also obtained in the genomes of vespertilionid bat species closely related to *M. lucifugus* (Brandt's myotis, *Myotis brandtii* [45]; David's bat, *Myotis davidii* [46]; big brown bat, *Eptesicus fuscus*).

Further examination revealed that the hits in the feline genomes corresponded to an endogenous gammaretrovirus family initially described in the domestic cat. Two proviruses of this family were initially documented in cat as FERVmlu1 and FERVmlu2 [47]. In 2011, this ERV family was also reported in Repbase [48] as ERV1-1_Fca. In a more recent and more systematic inventory of ERVs in the cat genome [49], this family was designated as FcERV_γ6, a nomenclature we will adopt hereafter. Most recently, this family was identified as part of "lineage VII" by Mata et al. [34] who also reported the presence of closely related gammaretroviral elements in several wildcat species, including jaguar, puma, jaguarundi and tiger. To our knowledge, the related elements in the pangolin have not been previously characterized elsewhere. Hereafter we refer to this novel ERV family as MPERV1 for *Manis pentadactyla* ERV1 and deposited its

Table 1. Copy number of MLERV1 related proviruses in different species.

	Tiger	Cat	*E. fuscus*	*M. davidii*	*M. brandtii*	*M. lucifugus*	pangolin
Full-length ERV	55	88	3	48	51	204	2
Solo LTR	675	744	67	1042	948	1638	27
total	730	832	70	1090	999	1842	29

doi:10.1371/journal.ppat.1005279.t001

consensus sequence in Repbase. For simplicity, we refer to all the elements detected in vespertilionid bats as MLERV1 and all the elements in different felids as FcERV_γ6.

To determine the ERV copy number in each species, we used the LTR sequences to mask their corresponding genome assembly using the Repeatmasker program and parsed the positional output to infer the number of putative full-length proviruses (i.e. containing two LTRs) and solitary (solo) LTR (see Methods). The results of this analysis (Table 1) show that each species harbors a relatively small number of full-length proviruses (2–50) but often numerous solo LTRs (up to 1600 + in *M. lucifugus*). It should be noted that the vast majority of proviruses we inferred to be full-length (based on the occurrence of a pair of LTRs within 10 kb) contain sequencing/assembly gaps. Thus we cannot ascertain whether they contain all the coding domains of a complete provirus.

A retroviral CST event involving bat, cat and pangolin

To illustrate the exceptional level of sequence similarity among MLERV1, FcERV_γ6 and MPERV1, we generated nucleotide pairwise alignments of FcERV_γ6 and MLERV1 and of FcERV_γ6 and MPERV1 using the most closely related full-length proviruses from each family and performed a sliding window analysis of nucleotide identity across the two pairwise alignments (Fig 1A). As a comparison, we performed the same analysis for proviral sequences representative of HIV-1 (Group M subtype B) and its closest relative from the chimpanzee SIVcpz [50,51]. The results show that the two representatives of the MLERV1 and FcERV_γ6 families and two representatives of FcERV_γ6 and MPERV1 are highly similar throughout their entire length, with an average level of nucleotide identity (~85%) comparable to that between HIV-1 and SIVcpz (Fig 1A). The most divergent segment corresponds to the predicted surface (SU) domain of the envelope protein (~50% identity in the N-terminal region). Elevated divergence in the SU region is also apparent between the two lentiviruses, as previously documented [52], and is thought to reflect the rapid adaptation of retroviral envelope to diverged host cell receptors [53]. In summary, MLERV1, FcERV_γ6 and MPERV1 are just as closely related to each other as HIV-1 and SIVcpz, and thus these three elements and their relatives in the bat, cat and pangolin genomes can be considered as endogenous elements descended from the same retrovirus.

The overall level of nucleotide similarity between MLERV1, FcERV_γ6 and MPERV1 is strongly incongruent with a scenario of vertical inheritance of an ancestral ERV present in the common ancestor of chiropterans, felids and pangolins, which dates back to ~85 million year ago (MYA) [54,55]. Furthermore, we could not find any close relative of MLERV1 or FcERV_γ6 (no megaBLAST hit with sequence identity >80%) in the genome assemblies of species representative of other chiropteran (e.g. flying fox, Pteropodidae) or carnivore families (e.g. dog, Canidae; bear, Ursidae; ferret, Mustelidae; seal, Phocidae; walrus, Odobenidae). The Chinese pangolin genome is the only available representative of the order Pholidota, which is considered sister to Carnivora, and thus equally related to Perissodactyla (horse, rhino) and Cetartiodactyla (cow, pig, hippo, whales), all of which appear to lack related ERVs (Fig 1B). Thus, the taxonomic distribution of MLERV1/FcERV_γ6 elements is extremely patchy, being detected in four vespertilionid bats, two feline species (cat and tiger), and one pangolin, but not in any of the numerous phylogenetically intermediate species represented in the NCBI WGS
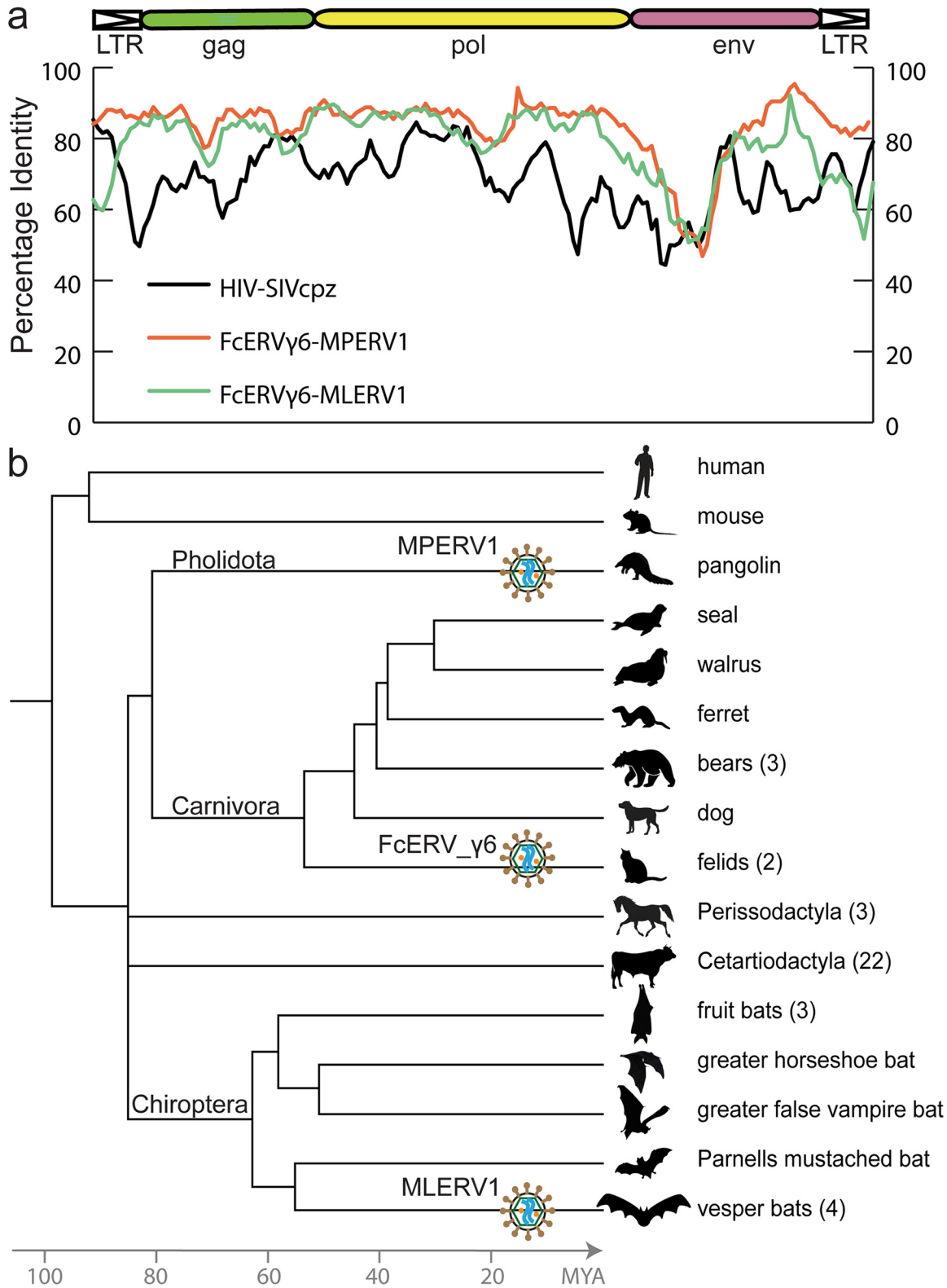
Fig 1. High sequence similarity and taxonomic distribution of MLERV1, FcERV_γ6 and MPERV1. (a) Sliding window analysis of percent sequence identity along pairwise alignments of entire proviruses. DNA sequence distance is corrected using kimura 2 parameter substitution model. (b) Taxonomic distribution of MLERV1, FcERV_γ6 and MPERV1. A schematic of the phylogenetic relationship of the 55 species from the clade "Scrotifera" currently

represented in the NCBI whole genome sequence database, with human and mouse shown as outgroups. The 55 species fall within 6 mammal orders: Pholidota, Carnivora, Cetacea, Artiodactyla, Perissodactyla, Chiroptera. Some of the species are collapsed by order/family with the number of species for each clade indicated into parentheses. The three independent retroviral invasions of MLERV1, FcERV_γ6 and MPERV1 are depicted above each of the mammal lineages affected. The placement of retroviral particles does not imply the timing of corresponding CST.

database (Fig 1B). This taxonomic distribution suggests that the retrovirus that gave rise to MLERV1, FcERV_γ6 and MPERV1 underwent at least two CST events and was endogenized at least 3 times independently in the vespertilionid, felid, and pangolin lineages.

Dating MLERV1/FcERV_γ6 insertions using comparative genomics

To gain further insights into the evolutionary history of these ERVs, we next sought to estimate when they first infiltrated their host genomes. Given that the likelihood of the same endogenous retrovirus to integrate at the same exact genomic location independently in different lineages is negligible, the presence of an element at orthologous position in different species can be interpreted as having inserted prior to their divergence time [56,57]. Conversely, since ERVs are not known to excise from the genome, the absence of an element in one species at a genomic location occupied by an ERV in another species strongly suggests that the ERV integrated after the split of the two species [58,59]. Such 'empty' sites can be corroborated by the presence of a single copy of the host target sequence duplicated upon proviral integration (typically 4-bp target site duplication for gammaretroviruses). This cross-species presence/absence approach has been widely applied to date a variety of mobile element insertions, including ERVs [14,42,59,60]. It is possible that the age of some integration events may be underestimated because of incomplete lineage sorting. Therefore, orthologous insertion analysis should be interpreted with caution when applied to rapidly radiating species such as the three *Myotis* considered here.

We first examined the sharing of FcERV_γ6 elements between the cat and tiger, which diverged ~10.8 MYA [61]. Out of a total of 1,419 putative full length proviruses and solo LTRs detected in the current whole genome assemblies of the two species, we were able to ascertain that 256 occupy orthologous positions, while 261 and 201 are specific to the cat and tiger lineages, respectively. None of these elements were detectable in other available carnivore genome assemblies (e.g. dog, panda, ferret, seal), while some of their flanking host sequences were readily detected (e.g., the flanking sequence of FcERV_γ6–68 is found in dog chromosome 14). These data indicate that FcERV_γ6 first invaded a felid ancestor sometime between ~10.8 million years (MY) and ~55 MYA and has continued to amplify to generate many insertions specific to the cat and tiger lineages (Fig 2).

Our previous phylogenetic analysis [42] has shown that the MLERV1 family of the little brown bat *M. lucifugus* can be divided into 3 subfamilies. Here we performed a systematic analysis of the presence/absence of MLERV1 elements (including solo LTRs) from the 3 subfamilies in the genome assemblies of three other vespertilionid bats currently available: Brandt's myotis (*Myotis brandtii*), David's bat (*Myotis davidii*) and the big brown bat (*Eptesicus fuscus*), which have been estimated to diverge from *M. lucifugus* ~10 MYA, ~13 MYA and ~25 MYA, respectively [62–65]. The vast majority of MLERV1 elements and their close relatives were found to be species-specific (Fig 2). Only 3 elements were present at orthologus loci across the 3 *Myotis* genomes (Fig 2) and we could not find a single insertion shared between *E. fuscus* and any of the *Myotis*. Another interesting observation is that members of the MLERV1_3 subfamily, which contributes the vast majority (>80%) of MLERV1 elements in the 3 *Myotis* genomes, could not be identified at all in the *E. fuscus* genome. Indeed, all 29 elements detected in *E. fuscus* cluster with either one of the other two subfamilies (S6 Fig). Together these data suggest

that the MLERV1 family expanded independently in the *Myotis* and *Eptesicus* lineages, but achieved a much higher copy number in the *Myotis* lineage due to the amplification of the MLERV1_3 subfamily, which has generated numerous species-specific insertions (Fig 2).

Dating of individual provirus insertions using LTR-LTR divergence

Another widely applied method to date retroviral and other LTR-bearing retroelement insertions relies on the divergence of the 5' and 3' LTR of individual elements. This is because their retrotransposition mechanism results in two identical LTRs at the time of chromosomal integration. Given that most ERV LTR sequences are assumed to evolve neutrally once integrated in the host chromosome, the age of a provirus can be estimated based on LTR divergence by applying the host neutral substitution rate [49,59,66,67]. To eliminate the inflated divergence caused by hypermutable methylated CpG sites [68], we excluded all the CpG sites from our calculation of LTR-LTR divergence. We applied this method to calculate the age of all complete (i.e. with two LTR) proviruses detected in cat, *M. lucifugus* and pangolin genome assemblies. We use previously estimated neutral substitution rates of 2.7×10^{-9} and 1.8×10^{-9} per year for vespertilionid bats and felids respectively [44,69], and an "average" mammal neutral substitution rate of 2.2×10^{-9} per year [70] for the pangolin.

The results of these calculations predict that the oldest MLERV1 and FcERV_γ6 proviruses would be ~10 MY and ~20 MY respectively (Fig 3A). The amplification of the bat MLERV1 family would have peaked sharply in the last 2 MY, while the cat FcERV_γ6 elements inserted more continuously over the past ~15 MY (Fig 3A). The two MPERV1 proviruses identified in the pangolin genome are estimated to be ~10 and ~18 MY based on this approach.
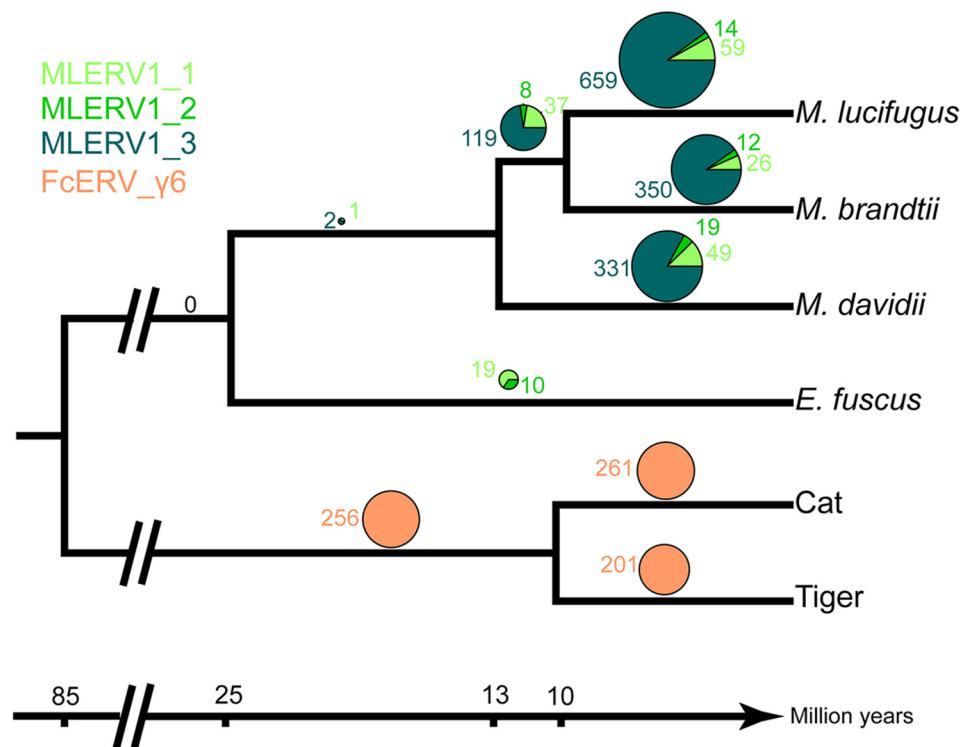

Fig 2. **Distribution of MLERV1/FcERV_γ6 insertions in vesper bats and felids.** The numbers of ERV insertions detected as orthologous or species-specific are shown as pies above each branch of the phylogeny of the vesper bats and felids examined. Different colors are used to illustrate FcERV_γ6 and the three different MLERV1 subfamilies.
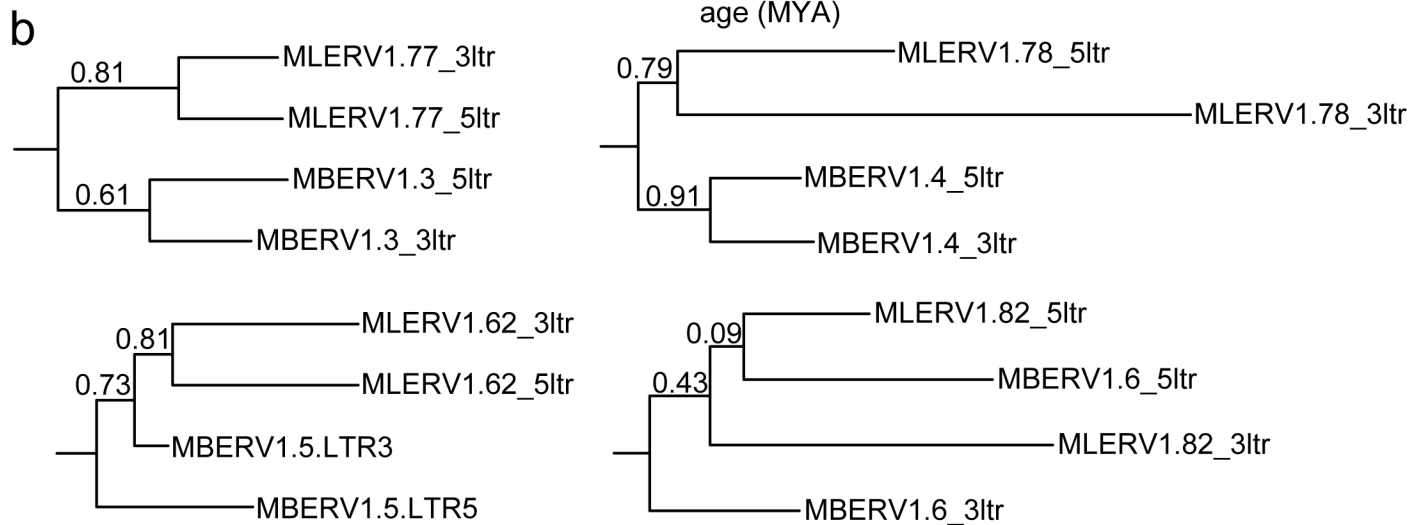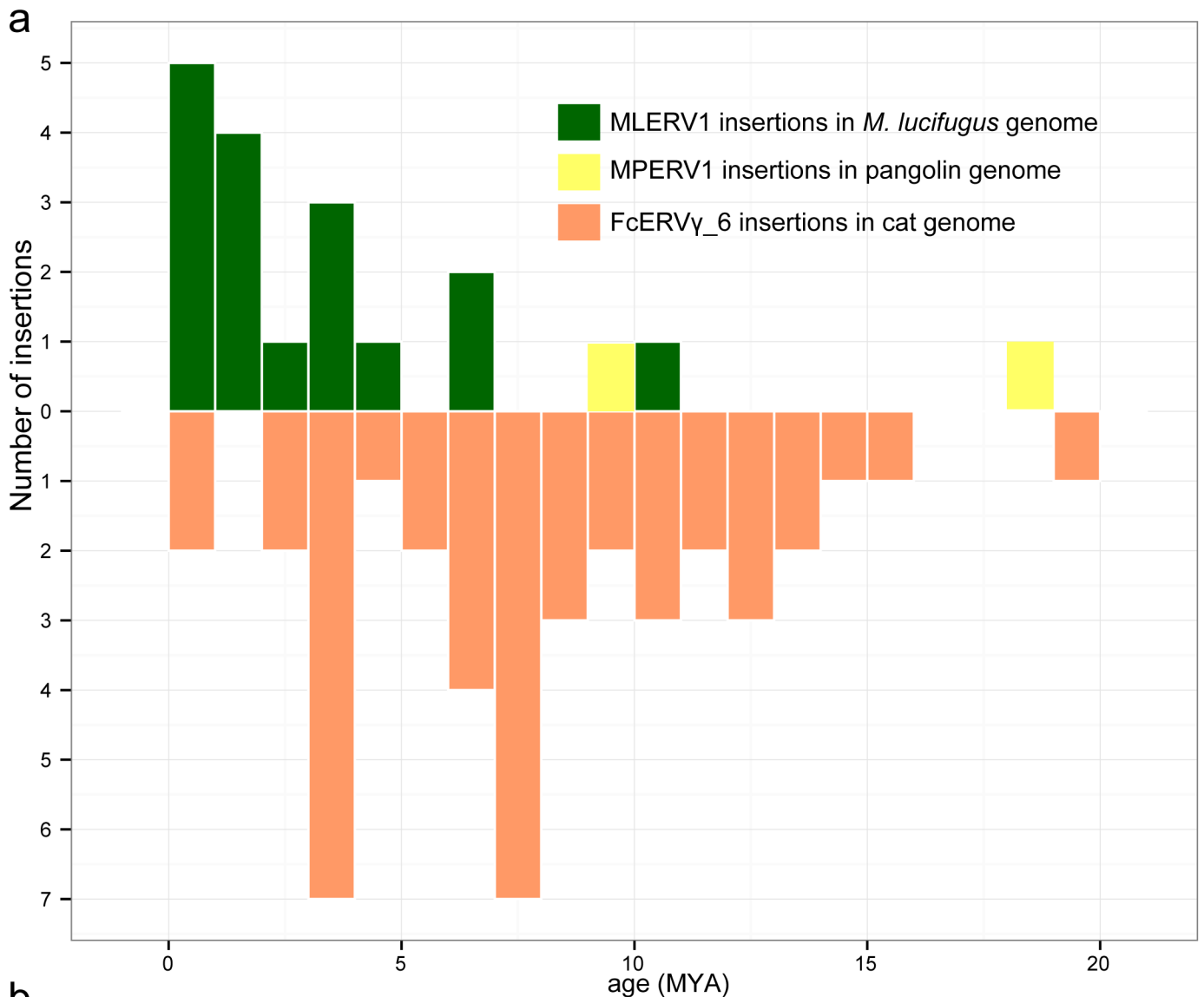
doi:10.1371/journal.ppat.1005279.g002

Fig 3. Dating individual proviral insertions based on LTR-LTR divergence. (a) Age distribution of proviral insertions inferred from LTR-LTR divergence. The *y* axis shows the number of insertions for each age class binned in MY on the *x* axis. Each ERV family is shown as bars of different colors. (b) Evidence of 'gene' conversion between 5' and 3' LTR of the same provirus. Four LTR trees are shown for four pairs of orthologous proviruses shared by *M. lucifugus* (MLERV) and *M. brandtii* (MBERV). Each maximum likelihood tree was built from a multiple alignment of the 5' and 3' LTRs from each provirus rooted with a non-orthologous LTR from *M. lucifugus* (also illustrated in S2 Fig). The support for each node as determined with an approximate likelihood ratio test (aLRT) is shown. The fact that 5' and 3' LTR from the same provirus tend to group together rather than by species is indicative of gene conversion between the LTRs along the two species lineages following proviral insertion in their common ancestor.

While these estimates are consistent with independent ERV invasions of the vespertilionid, felid and pangolin lineages, we noticed that the age of individual insertions based on LTR divergence were generally lower than those estimated based on their presence/absence at orthologous position across species. For instance, we found that 27 FcERV_γ6 proviruses were orthologous in cat and tiger, which indicates that all must have inserted prior to speciation of these felids, which has been robustly estimated at ~10.8 (8.4–14.5) MY [61]. However only 13 of these 27 insertions were estimated to be older than 10 MY based on LTR divergence (S1 Table). Similarly, *M. lucifugus* and *M. brandtii* are thought to have diverged ~10 MYA [62,64], but the age of the four MLERV1 insertions orthologous between these two species was estimated to be 10.5, 6.8, 4.5 and 1.2 MY based on LTR divergence (S1 Table). One possible explanation for these discrepancies between the two dating methods is the phenomenon of gene conversion between two LTRs adjacent in the genome, which essentially erases some of the divergence accumulated over time through point mutations occurring in each of the LTRs, causing to underestimate the date of proviral insertion [71,72]. Indeed, a phylogenetic analysis of the LTR sequences from the four MLERV1 proviruses orthologous in *M. lucifugus* and *M. brandtii*, shows topologies consistent with LTR homogenization through gene conversion events for at least two of the proviruses examined (corresponding to the two upper trees in Fig 3B): their 5' and 3' LTR cluster together rather than by species (Fig 3B). This is the topology predicted if conversion events in one or both of the species lineages had removed nucleotide divergence accumulated between 5' and 3' LTR prior to speciation [72]. Thus, estimates of the age of proviruses based on LTR divergence should be interpreted with caution, as they are likely to be underestimates. Nonetheless, the results are in agreement with the other lines of evidence that the vespertilionid, felid, and pangolin lineages were independently infiltrated by the same ERV during an evolutionary timeframe ranging from ~25 to ~13 MYA.

Phylogenetic analysis of FcERV_γ6, MLERV1 and MPERV1 families

To further characterize the evolution history of MLERV1, FcERV_γ6 and MPERV1, we examined the phylogenetic relationship of elements within these families using a maximum-likelihood tree built from an alignment of their 3' LTR sequences. We used only 'complete' proviruses (30 in bats, 43 in cats and 2 in pangolin), since we observed that including tiger provirus and solo LTRs did not yield any new major clade in the phylogeny (S1 Fig). Also the general topology of the tree is identical if the 5' LTR sequences are included (S2 Fig). Trees generated using internal coding sequences also displayed the same general topology (S3 Fig), but offered less phylogenetic resolution due to the more constrained nature of retroviral coding sequences relative to LTRs [73,74].

The unrooted tree resulting from the phylogenetic analysis (Fig 4) clearly shows that FcERV_γ6 and MPERV1 elements are more closely related to each other than to the bat MLERV1 elements. Another striking observation is that elements within the FcERV_γ6 family fall within a single clade with uniformly short branches, whereas the MLERV1 elements, as we previously reported [42] can be divided into 3 distinct subfamilies separated by long branches, with MLERV1_2 and MLERV1_3 being closer to each other and more distant from FcERV_γ6
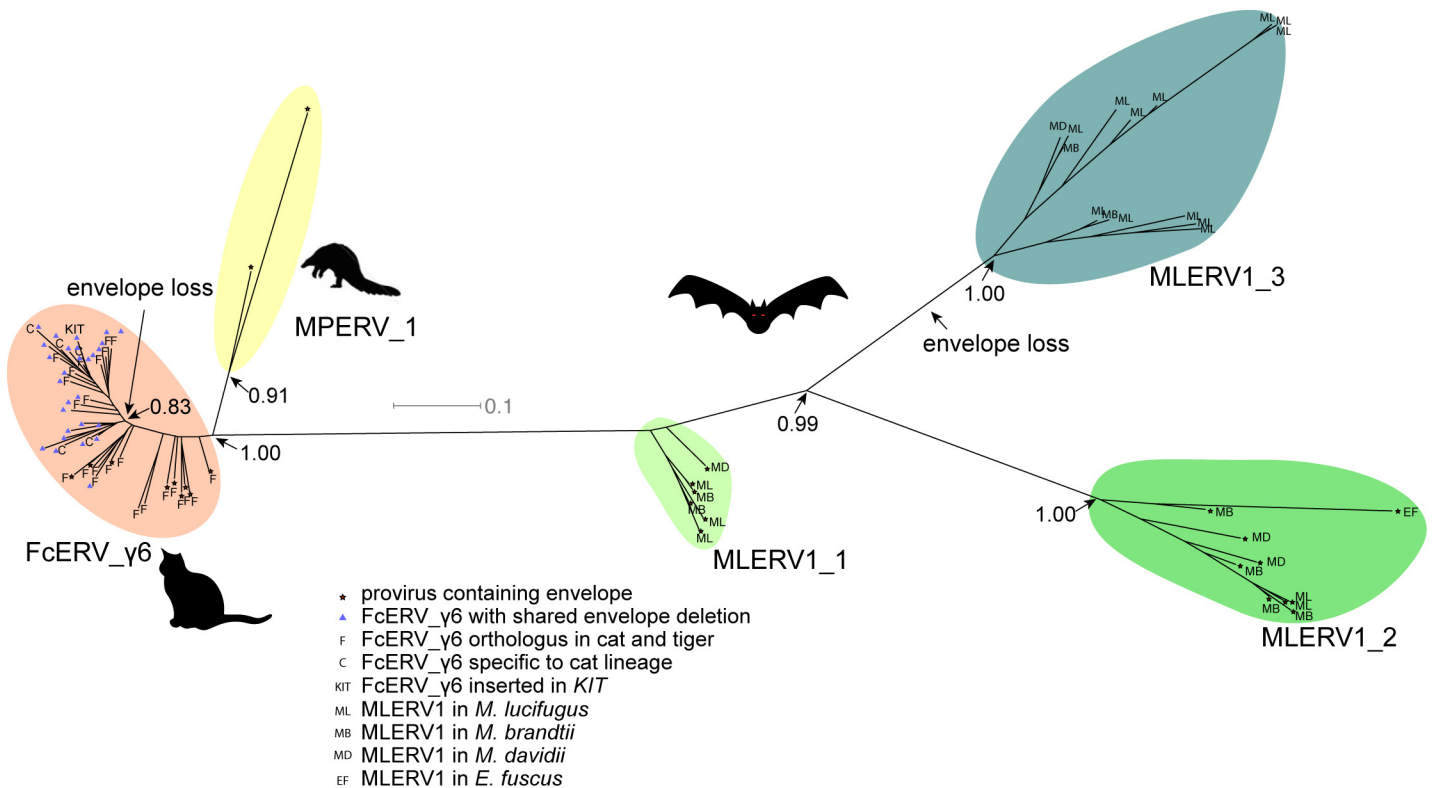
Fig 4. Phylogenetic analysis of MLERV1, FcERV_γ6, MPERV1 families. A maximum likelihood phylogenetic tree built from a multiple alignment of 3' LTR sequences of 75 proviruses. The support for each node as determined with an approximate likelihood ratio test is shown. Information on the species origin as well as the presence/absence of envelope sequence is labeled at each node. Two independent losses of envelope by deletion are highlighted.

than MLERV1_1 (Fig 4). These data are consistent with a scenario whereby the FcERV_γ6 family was amplified from a single infectious progenitor, while MLERV1 elements might have originated from at least three distinct infectious progenitors.

Selection analysis on coding sequences reveal different amplification dynamics

To further explore the history of the FcERV_γ6 and MLERV1 families, we next turn to an analysis of selection regimes that have acted on their coding sequences during their amplification. Such analysis can help discern whether ERVs have spread primarily through reinfection or retrotransposition events because the latter mechanism, which is strictly intracellular, is predicted to be associated with the loss of envelope function. Indeed, the envelope protein binds to host cell membrane receptor to promote virion entry in the host cell and therefore is required for most retroviral infection [5]. Thus, proviruses that originate from infection events should show evidence of functional constraint on envelope domains [75,76], Magiorkinis:2012gy}. To perform this analysis, we used all MLERV1 (n = 30) and cat FcERV_γ6 (n = 43) proviruses with complete (or nearly complete) coding capacity. Given that only 2 proviral MPERV1 copies could be identified, we did not perform selection analysis for this family.

To evaluate how natural selection may have constrained the different coding regions of MLERV1 and FcERV_γ6, we computed the dN/dS ratio (ω) applying the branch model implemented in PAML, where dN denotes the non-synonymous substitution rate and dS denotes the synonymous substitution rate, along the branches of the phylogeny of FcERV_γ6 and
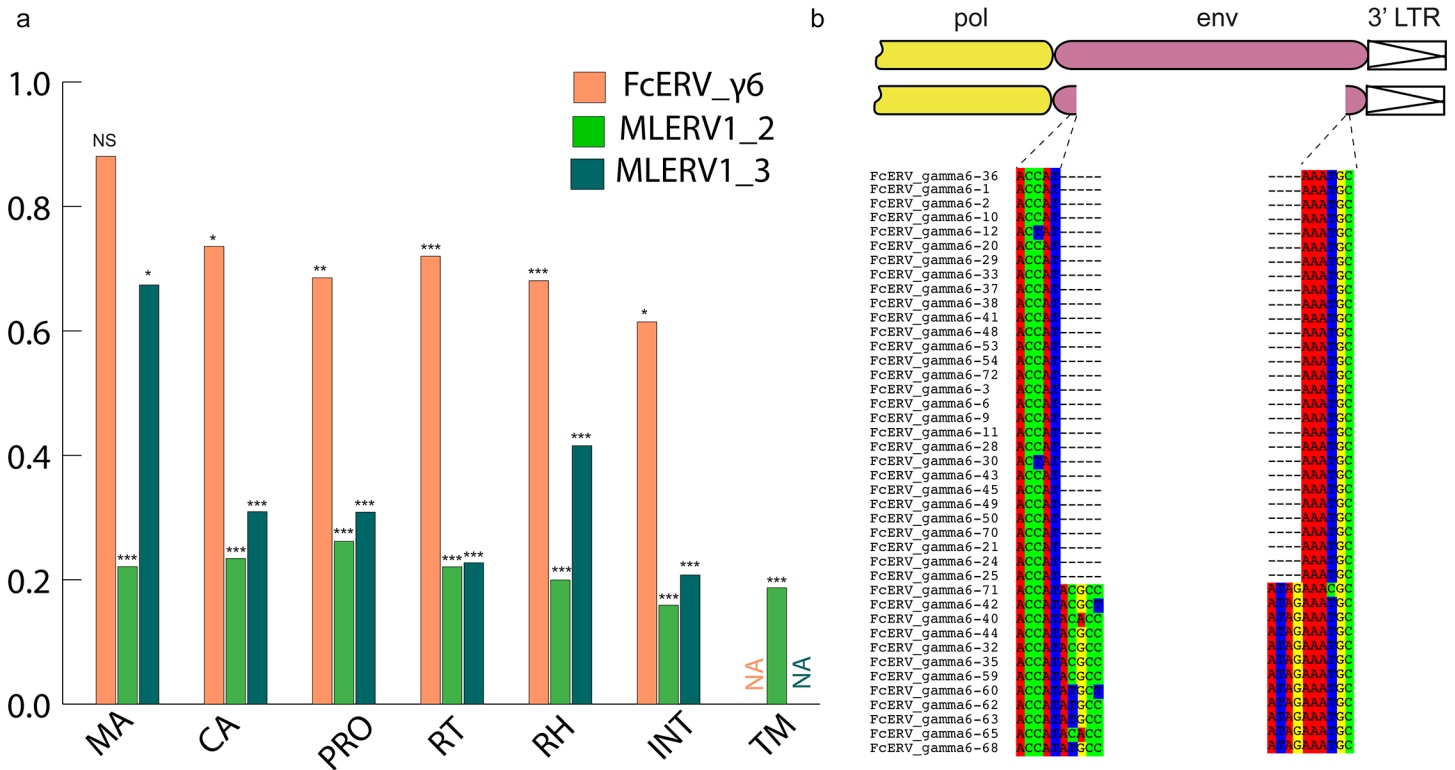
Fig 5. Selection analysis on coding domains. (a) dN/dS ratio (ω) of each coding domain in FcERV_γ6, MLERV1_2 and MLERV1_3. MA, CA, PRO, RT, RH, INT and TM denote matrix, capsid, aspartyl protease, reverse transcriptase, RnaseH, integrase and envelope transmembrane domain, respectively. Asterisks denote the level of significance of departure from ω = 1 (likelihood ratio test, see Methods) with * = p<0.05; ** = p<0.01; *** = p<0.001. NS = not significant (p>0.05); NA = not applicable (domain deleted). (b) Shared breakpoints at the site of envelope deletion in a subset of FcERV_γ6 elements. A schematic of the prototypical proviral coding regions showing the approximate position of the envelope deletion in 29 FcERV_γ6 elements marked with blue triangles in Fig 4 and (below) an alignment with a subset of envelope-containing FcERV_γ6 elements, showing that they share the same deletion breakpoints. These data indicate that these 29 elements likely arose from amplification of a progenitor copy that had suffered a large deletion in the envelope region.

doi:10.1371/journal.ppat.1005279.g005

MLERV1 elements for each of their predicted coding domains [77]. ω values significantly smaller than 1 are indicative of purifying selection acting to maintain a functional protein sequence, while ω values not significantly different from 1 are indicative of neutral evolution or relaxed functional constraint. To test for significant deviation of ω from 1, we apply a likelihood ratio test [78].

Within the FcERV_γ6 family, the analysis reveals that purifying selection has acted on all coding domains (ω value ranging from ~0.6 to ~0.9, $p < 0.05$), with the notable exception of Gag matrix and envelope domains (Fig 5A). The ω value is not significantly different from 1 (neutral evolution) for the Gag matrix domain (Gag_MA). Besides, all but nine of the 43 FcERV_γ6 proviruses lack an envelope domain (TLV_coat). The nine copies that have retained a recognizable envelope domain occupy basal branches in the phylogeny (Fig 4) and have orthologs in the tiger genome suggesting that they predate the envelope-less copies (S1 Table). Furthermore, 29 of the 43 FcERV_γ6 proviruses examined, including all cat-specific copies, share the same deletion breakpoint removing most of envelope gene (Fig 5B). These data suggest that FcERV_γ6 copies potentially coding an envelope were inserted prior to the speciation of cat and tiger (~10.8 MYA), while copies integrated more recently lacked the envelope domain. In addition, the envelope open reading frames of these nine ancient FcERV_γ6 elements accumulated multiple indels or missense substitutions. Thus, none of the FcERV_γ6

elements in the cat genome appear to have retained a functional envelope domain. These data suggest that FcERV_γ6 rapidly lost its infectious capacity in the cat lineage but has continued to amplify primarily via retrotransposition amplified primarily via retrotransposition.

By contrast, selection analysis suggests that the MLERV1 family has experienced a more complex amplification history. We focused our analysis on the MLERV1_2 and MLERV1_3 subfamilies because they are the two best-supported monophyletic subfamilies with sufficient number of proviruses to draw solid conclusions. First, we observe that generally the signature of purifying selection is more pronounced on the bat elements than on the cat elements, as indicated by much lower ω values (Fig 5A). The only exception is the Gag matrix domain of the MLERV1_3 subfamily, which exhibits relatively higher ω value (ω = 0.67, p = 0.02) (Fig 5A). In addition, all MLERV1_3 elements appear to have lost their envelope domain through the same deletion event (Fig 4). This pattern contrasts with elements within the MLERV1_2 subfamily, for which all coding domains, including envelope, have evolved under strong purifying selection during the spread of these elements (ω from 0.15 to 0.27, p<0.001) (Fig 5A). These data suggest a scenario whereby MLERV1_3 has amplified primarily by retrotransposition, while the spread of MLERV1_2 has been driven by multiple infection events.

We also observe that in both FcERV_γ6 and MLERV1_3, the losses of envelope coincided with the elevation of the dN/dS ratio in their Gag matrix domain (Fig 5A). To evaluate whether this reflects a loss of function (neutral evolution) or a relaxation of purifying selection, we further examined the integrity of open reading frames (ORFs) in each ERV family by computing the frequency of stop codons and frameshift mutations occurring in each of the domains (see Methods). Overall the results indicate that the coding integrity of the Gag matrix domains of FcERV_γ6 and MLERV1_3 elements is not significantly different from that of the other ERV subfamilies or that of the other protein domains (S4 Fig). These results suggest that the Gag matrix domain is not dispensable for retrotransposition, as previously demonstrated functionally for IAP elements [79], but appears to evolve faster in retrotransposing ERVs.

Discussion

Cross-ordinal transmission of a mammalian retrovirus

Until recently, most retroviral CST events that have been documented rigorously have implicated closely related species [6,9], suggesting that the phylogenetic distance between species is an important determinant of the host range of a retrovirus [7,29]. Indeed, many previous studies have illustrated how the divergence of host cellular factors that either facilitate or restrict viral replication can modulate the host range of a virus [80–82]. The systematic analysis of retroviruses fossilized in the genome as ERVs is progressively revealing a more nuanced picture whereby some retroviruses appear to have been capable to infect widely diverged species (i.e. belonging to different orders) without seemingly much changes occurring in their own sequences. Recent large-scale pylogenomics analyses have suggested that cross-order transmission may actually be fairly common for some groups of retroviruses, including gammaretroviruses [19,20,34] and IAP betaretroviruses [33]. While these studies disclosed phylogenetic patterns suggestive of multiple CST events, they did not explicitly rule out alternative hypotheses, such as vertical persistence and stochastic loss of the ERV in some lineages, and thus they generally await confirmation through more detail analyses such as the one presented here.

Our study provides multiple lines of evidence supporting the notion that a gammaretrovirus infiltrated independently the germline of bat, cat and pangolin species representing three mammalian orders (Chiroptera, Carnivora, Pholidota, respectively). First, elements found in these species display a level of nucleotide sequence similarity (~85%) along their entire length that is comparable to that observed between closely related retroviruses that have undergone very

recent CST (such as SIVcpz and HIV-1). Such a level of sequence similarity between ERVs inhabiting species diverged by ~85 MYA [54] is incompatible with a scenario of vertical descent from an ERV inherited from their common ancestor. The CST hypothesis is also bolstered by the highly discontinuous taxonomic distribution of this particular ERV family. Out of 107 mammal species for which whole genome assemblies are publicly available, we could only detect members of this ERV family in vespertilionid bats, felids and pangolin, but not in several species representing related mammal families (6 additional Chiroptera species from 4 families and 7 additional Carnivora species from 5 families). Thus, a scenario evoking a single introduction of this ERV family in the common ancestor of bats, cats and pangolins followed by vertical inheritance would necessitate at least 5 independent losses (Fig 1B) to account for its current taxonomic distribution. A more parsimonious scenario is that this ERV family was acquired horizontally and independently in each of the three species lineages where it is currently detected. It is also possible that the ancestral retrovirus infected other species but failed to endogenize in their genomes, or it could also be that additional species lineages hosted this ERV family but have gone extinct. Finally, our estimation of the dates at which these elements first entered their host genomes, which relies on two independent approaches (cross-species comparison of orthologous ERV loci and LTR-LTR divergence), converges to a bracket of 13 to 25 MYA, which far postdates the divergence of their host species (~85 MYA). Together these data indicate that a progenitor gammaretrovirus infiltrated the germlines of ancestral vespertilionid bat, felid and pangolin species.

It is conceivable that this retrovirus could have transferred directly between these ancestral species because their geographic distribution likely overlapped in Eurasia during the estimated period of initial ERV infiltration (~13–25 MYA) [61,62,83]. Given that cats are known to prey on both bats [84–87] and pangolins [88], a direct transfer from bat or pangolin to cat is plausible. Indeed, predation has been put forward as the most likely explanation for the spillover of bat lyssaviruses (rabies) into domestic cats [89]. On the other end, both bats and pangolins are capable of surviving a cat attack, which makes the transfer from predator to prey conceivable as well. Nonetheless, multiple lines of evidence indicate that MLERV1 colonized these bat genomes more recently (Figs 2 and 3), which may suggest a CST from cat to bat. Furthermore, we cannot rule out that one or several intermediate hosts were involved in the introduction of the retrovirus in these species.

Repeated transition from retrovirus to retrotransposon

Our data suggest that, shortly after infiltration of the felid genome, FcERV_γ6 lost the capacity to infect cells and transformed into a retrotransposon. Envelope domain remnants are only found in the basal branches in their phylogeny, and all the FcERV_γ6 elements amplified in the domestic cat lineage clearly derive from a progenitor that lacked coding capacity for a functional envelope protein (Figs 4 and 5B). Together these data suggest that FcERV_γ6 lost its infectious capability soon after it became endogenous, but continued to propagate by retrotransposition, much like the IAP elements in the mouse genome [90,91]. Coincided with the envelope loss, we found gag matrix domain evolves at a relaxed rate in FcERV_γ6 family. A recent study showed that a FcERV_γ6 insertion in the *KIT* gene currently segregating in domestic cats is responsible for the "Dominant White" and white spotting pigmentation phenotypes [92], which supports our findings that some FcERV_γ6 insertion activity is very recent and likely ongoing. Interestingly, the FcERV_γ6 element inserted at the *KIT* locus lacks envelope domain and clusters with other recently active FcERV_γ6 copies in our phylogenetic analysis (Fig 4). Collectively these data suggest that FcERV_γ6 has morphed into a successful retrotransposon that may still be active in the domestic cat.

In contrast to FcERV_γ6, the sequence diversity and phylogenetic structure of MLERV1 elements in the vesper bat genomes are indicative of a more complex amplification history characterized by a mixture of retrotransposition and reinfection events. Our phylogenetic analysis delineates at least three highly diverged MLERV1 subfamilies. The separation between three subfamilies suggested that this family stemmed from at least three related infectious progenitors independently, which is conceivable considering the independent introduction of multiple HIV-1 strains in human population [93]. Furthermore, selection analyses suggest that different subfamilies have adopted different evolutionary trajectories. The MLERV1_2 subfamily is characterized by a signature of intense purifying selection acting on all coding regions throughout the whole clade (Fig 5A). These data strongly suggest that elements within that subfamily have retained their infectious capacities for extended period of time and most likely spread primarily through reinfection events. It is even possible that MLERV1_2 is still active and infectious: most insertions are very recent (Fig 2 and S1 Table) and at least one copy (MLERV1.80) contains apparently full-length and intact *gag*, *pol*, and *env* genes.

The MLERV1_3 subfamily appears to have followed a different evolutionary path whereby the divergence of the elements was accompanied by a strong signature of purifying selection in all coding regions with the notable exception of the Gag matrix domain which has been evolving faster than other domains (Fig 5A) and the envelope domain which was apparently deleted altogether. This selection pattern resembles that of FcERV_γ6 and is indicative of proliferation primarily via retrotransposition as opposed to reinfection. Consistent with this hypothesis and the so-called superspreader model [33], the MLERV1_3 family has been by far the most successful at spreading during *Myotis* evolution: it has the highest copy number, including many species-specific insertions (Table 1).

Interestingly, none of the 29 MLERV1 elements identified in the big brown bat *E. fuscus* belong to the MLERV1_3 subfamily. This is consistent with the idea that the MLERV1_3 subfamily originated after the split of Eptesicus-Myotis split ~25 MYA and amplified during the diversification of the *Myotis* lineage. At present it remains unclear whether the MLERV1 elements present in *E. fuscus* and *Myotis* descend from element(s) introduced in their common ancestor or if they result from independent acquisition of the same retrovirus. On the one hand, the observation that both *E. fuscus* and *Myotis* harbor elements from two diverged subfamilies may be interpreted as evidence that these subfamilies descend from a single progenitor ERV acquired in the common ancestor of these species. On the other hand, the fact that none of the MLERV1 insertions are shared (orthologous) between *E. fuscus* and any of the 3 *Myotis* genomes (Fig 2) and that none of the provirus insertions dated in any of these bat species appear older than 13 MY (considerably less than the estimated divergence between the two genera, 25 MYA) (Fig 3) supports a scenario of multiple, independent acquisition. This scenario, while requiring at least two CST events, is conceivable because *Eptesicus* and *Myotis* bats likely occupied a widely overlapping geographic distribution at the estimated time of MLERV1 invasions [62] and these congeners are currently known to frequently come into contact within the same roost [94,95].

Regardless of the origin of MLERV1, the data summarized above illustrate how the same retrovirus has infiltrated widely diverged mammals and transitioned multiple times (at least twice: FcERV_γ6 and MLERV1_3) from an infectious pathogen to a genomic parasite (i.e. a retrotransposon). The biological factors and sequence of events underlying such transition remain poorly understood. In a seminal study, Ribet et al. showed that the loss of envelope gene combined to the gain of an endoplasmic reticulum targeting signal were apparently sufficient for an infectious progenitor of the mouse IAP elements to turn into a highly active retrotransposon [79]. Magiorkinis et al. [33] have extended this paradigm and proposed that the passive loss of envelope lead ERVs to become "superspreaders" in the genome. Through a study of IAP-like elements across a wide range of species, these authors observed that

envelope-less elements generally achieve much higher copy numbers than those maintaining a functional envelope. Our results support this model. First, envelope-less FcERV_γ6 elements have proliferated to high copy numbers in the domestic cat (n = 832) and tiger (n = 730). In addition, in the bats the only subfamily of MLERV1 elements that has attained similarly high copy number is MLERV1_3, which conspicuously lack a functional envelope gene ([Fig 5](#)). MLERV1_3 elements have generated many species-specific insertions consistently outnumbering the MLERV1_2 subfamily, which appears to have spread primarily by reinfection (659 vs. 14 in *M. lucifugus* genome, 350 vs. 12 in *M. brandtii* and 331 vs. 19 in *M. davidii*) ([Fig 2](#)). Thus, our study is consistent with the notion that the loss of infectious capacity correlates with ERV expansion by retrotransposition, as proposed previously for the rodent IAP families [33,79].

One important difference is that the shift between infection and retrotransposition in the MERV1 family was apparently accompanied by little changes in the sequence of MLERV1 elements. Indeed, members of the MLERV1_2 and MLERV1_3 subfamilies diverge by ~15–20% in their RT domain nucleotide sequences with Kimura correction. By comparison, infecting and retrotransposing IAP subfamilies diverge more substantially (~65% in RT domain). Thus our findings suggest that the transition between the two modes of ERV amplification can occur relatively fast during ERV evolution.

Does host biology affect ERV proliferation?

An intriguing finding of this study is that the same or very similar retrovirus was endogenized in three different mammalian hosts, but followed quite different evolutionary trajectories in the three species lineages. In the pangolin lineage, the ERV family failed to amplify (only 2 detectable copies) and was essentially 'dead-on-arrival'. In the cat lineage, the ERV progenitor apparently lost its infectious capacity shortly after endogenization and subsequently amplified to high copy numbers by retrotransposition through an extended period of time ranging from at least 10 MYA (256 insertions orthologous in cat and tiger) to modern times (*KIT* insertion segregating in domestic cats). Meanwhile, in the bat lineage, the ERV followed a more complex evolutionary path characterized by multiple episodes of reinfection, and at least one burst of amplification by retrotransposition. These observations beg the question whether the loss of infectious capacity of an ERV and its conversion to a retrotransposon is a purely stochastic process, largely owing to the stochastic mutation of gag matrix and loss of envelope functions, or possibly the characteristics of different proviral ancestors, or if it can be influenced by some biological characteristics of the host species? For instance, it has been recently reported that the level of endogenous retroviral activity may be partly governed by host body size [96]. The pattern of sustained reinfection of MLERV1 in the bat lineage is particularly intriguing in light of the growing appreciation that bats seem to frequently act as reservoir for viruses otherwise lethal to other mammals [35,39]. The reasons for bats' propensity to support high and diverse loads of viral pathogens are poorly understood, but it is thought that some physiological (e.g. immunopathological tolerance) and/or ecological features (e.g. flight, roosting) allow these animals to tolerate higher level of viral replication and/or facilitate viral transmission [39,97,98]. By the same token, it is tempting to speculate that the same properties might predispose bats to support higher level of ERV reinfection compared to other mammals such as cats. However, we only investigated one ERV family in three mammal species here. It is possible that MLERV1 is just a unique ERV family in bat genome, and overall the ERV replication in bats is not significantly different from other mammals. Testing this hypothesis will necessitate a more systematic examination of the amplification dynamics of ERVs in a wide range of mammals to assess whether the tendency toward maintenance of infectious capacity is a general trademark of bats or possibly other groups of mammals.

Methods

Initial detection of CST events involving MLERVs

Nucleotide sequences of all RVT_1 domains of previously identified MLERVs [1,42] were used as queries to search the whole genome sequence database from the National Center for Biotechnology Information (NCBI) using default MegaBLAST parameters [99]. An 80% similarity over 80% region was used as filter to exclude non-specific hits.

Identification of complete proviruses, putative full-length ERVs and solo LTRs

Complete MLERV1 and FcERV_γ6 proviruses in the *M. lucifugus* and cat genomes were collected from previous publications [42,49]. To ensure we only considered elements from these families, we only retained elements with 80% nucleotide similarity to another family member, a procedure which resulted in the exclusion of the FcERV_γ6_46 copy from the FcERV_γ6 family (S3 Fig).

To identify complete proviruses in other vesper bat genomes, the RVT_1 domain sequence of MLERV1.71 in *M. lucifugus* was used as query in blastn search of the *M. brandtii*, *M. davidii* and *E. fuscus* genome assemblies available in NCBI. In parallel, we applied LTRharvest [100] and LTRdigest [101] as described previously [42] to identify all putative proviruses in each of the three bat genome assemblies. We then used BEDTools to intersect the coordinates of RVT_1 domain blastn hits with that of the candidate proviruses [102]. All the candidate proviruses intersecting with a MLERV1 RVT_1 hit were 'manually' inspected to refine their termini and confirm their identity as members of the MLERV1 family.

To comprehensively retrieve all proviruses and solo LTRs related to the FcERV_γ6/ MLERV1/MPERV1 families in each of their respective genomes, we run RepeatMasker [103] with default setting and a custom repeat library with representitives from all MLERV1/ MPERV1/FcERV_γ6 subfamilies against each genome assembly. The RepeatMasker output was then parsed using script parse_RMout_count_solo_and_full.pl to produce bed files of all complete solo LTRs and full length ERVs. We define a complete solo LTR as a sequence matching the LTR with missing less than 150 bp at their 5' termini and missing less than 10 bp at their 3' termini. We identified elements as putative proviruses those delimited by two LTRs in the same orientation separated by 3 kb to 10 kb of intervening sequence. Manual inspection of a subset of putative proviruses identified by this approach confirmed that most contained typical ERV coding sequences, though frequently interrupted by large sequence/assembly gaps. The LTR libraries and PERL scripts used for these analyses have been deposited on Github (https://github.com/xzhuo/orthologusLTR.git).

In the pangolin genome, our initial MEGAblast search yielded only 2 significant hits to the MLERV1 RVT_1 domain, but many more related ERVs could be retrieved using the RT domain from these two initial hits in reiterative blast searches against the pangolin genome assembly. To examine the relationship of these RT elements to each other and to the MLERV1 and FcERV_γ6 families, we conducted a phylogenetic analysis using the Maximum Likelihood package PhyML3.1 with the GTR+Γ model [104]. The resulting tree (S3 Fig) revealed that only the two initial hits clustered with FcERV_γ6/MLERV1 and were considered part of the MPERV1 family. The other elements form a distinct family we called MPERV2. MPERV1 and MPERV2 elements share less than 80% nucleotide sequence similarity in their coding regions, but still retain substantial level of sequence similarity in their LTRs. Thus, to correctly estimate the number of solo LTRs for the MPERV1 family, we had to examine their position on a phylogenetic tree of LTRs (S5 Fig). Using this approach, 27 solo LTRs could be assigned to the MPERV1 family in the pangolin genome assembly (as reported in Table 1). Because MPERV2

was much more distantly related to FcERV_γ6 and MLERV1 (<80% sequence similarity), we did not analyze further MPERV2 in this study. Reference sequences for MPERV1 and MPERV2 have been deposited in Repbase [48].

All identified ERVs are available as bed format (S1 File).

Sliding window pairwise similarity calculation

To generate the sliding window analysis shown in Fig 1A, we used MUSCLE [105] to align the nucleotide sequence for three pairs of proviruses: MLERV1_77 vs. FcERV_γ6_62; FcERV_γ6_62 v.s MPERV1_ltr106; SIVcpz(AF115393) vs. HIV-1(NC_001802) and used SEA-VIEW [106] to manually adjust each alignment. Each pairwise alignment was then split into 300 bp windows with step size 50bp (i.e. = 170 segments for the MLERV1_77 vs. FcERV_γ6_62 alignment) and the percentage of sequence similarity was computed and corrected using kimura 2 parameter model for each window [107].

ERV orthologous loci identification

Orthologous ERV loci were detected similarly as we described previously [42]. Briefly, we used the Perl script extract_flanking_fasta.pl to extract 200 bp at both ends of each query element along with 200 bp of flanking sequences. The output file is then used as query in a batch blastn search against the target genome assembly with default parameters. The csv format blast output is then parsed using orthoblast_finder.pl to pair 5' end hits with 3'end hits. Finally, the paired hits output was parsed using the script final_annotation.pl to infer the presence/absence of each element in the target genome. All these perl scripts were deposited on Github (https://github.com/xzhuo/orthologusLTR.git).

Estimation of individual provirus insertions using LTR–LTR divergence

Sequence divergence between 5' and 3' LTRs from the same provirus was computed as previously described [42]. To infer insertion dates from LTR divergence of MLERV1 and FcERV_γ6 elements, we used previously estimated lineage-specific neutral substitution rates of 2.7×10^{-9} yr^{-1} [69] and 1.8×10^{-9} yr^{-1} [44] for the vespertilionid and felid lineages respectively. Since no substitution rate has yet been estimated for the pangolin lineage, we used the 'average' mammal neutral substitution rate of 2.2×10^{-9} yr^{-1} [70] to infer the age of MPERV1 insertions.

Phylogenetic analysis

The maximum-likelihood phylogenies presented for LTR sequences were built using PhyML3.1 and the support for each node is determined with an approximate likelihood ratio test [108]. The multiple alignment of LTR sequences was constructed using MUSCLE and PRANK with default nucleotide parameters and manually adjusted using SEAVIEW [105,106,109]. Nucleotide substitution model was chosen using AIC criterion in jmodeltest2.1.6 [110] (GTR + Γ). Dendroscope 3 was used for tree visualization [111].

Selection and integrity analysis on coding domains

ERV coding regions were predicted using HMMER3 in all 6 reading frames [112] to delineate Gag_MA (matrix), Gag_p30 (capsid), RVP (protease), RVT_1 (reverse transcriptase), RnaseH, rve (integrase) and TLV_coat (envelope transmembrane domain) domains in MLERV1, FcERV_γ6 and MPERV1 proviruses. A multiple codon alignment was generated for each set of coding domains using MUSCLE and manually adjusted with SEAVIEW [106]. The program codeml from the PAML4.8 package [77] was used to estimate dN/dS ratio with branch

model = 2. A maximum likelihood phylogeny of the LTR sequences was used as the guide tree in codeml. To test for purifying selection on each coding domain, we calculated the control lnL value by running codeml with ω fixed to 1. Then likelihood ratio test was performed as suggested by PAML to test if ω is significantly different from 1 [78].

Coding region integrity was assessed by calculating the frequency of stop codon or frameshift indels per codon. We calculated the total length of each domain in every subfamily, and counted the occurrence of stop codon and frameshift indels. The mean frequency and 95% confidence interval is calculated using the Poisson distribution [113].

Supporting Information

S1 Table. Excel table containing estimated age, ortholog and presence/absence of envelope of each provirus.
(XLSX)

S1 Fig. Unrooted ML tree of all related cat and tiger LTRs.
(PDF)

S2 Fig. The unrooted ML tree of all related proviruses in three lineages built with both 5' and 3' LTRs.
(PDF)

S3 Fig. Rooted RT domain ML tree of all related proviruses in three lineages with FcERV_γ6–46 used as outgroup.
(PDF)

S4 Fig. Integrity analysis of coding domains. Y axis represents frequency of stop codon or frameshift indel per codon in each domain, four different subfamilies are illustrated with different colors. Error bars represent 95% confidence interval.
(PDF)

S5 Fig. Pangolin soloLTR ML tree. All members of MPERV1 are enclosed within an arch, and the aLRT support of MPERV1 is labeled at the branch node.
(PDF)

S6 Fig. The ML tree with all *E. fuscus* soloLTRs and selected LTRs from other bats. Three subfamilies of MLERV1 are illustrated. SoloLTRs in *E. fuscus* cluster with either MLERV1_1 or MLERV1_2.
(PDF)

S1 File. Compressed bed files of all identified ERVs.
(ZIP)

Acknowledgments

Author Contributions

Conceived and designed the experiments: XZ CF. Performed the experiments: XZ. Analyzed the data: XZ CF. Contributed reagents/materials/analysis tools: XZ. Wrote the paper: XZ CF.

References

1. Parrish CR, Holmes EC, Morens DM, Park E-C, Burke DS, Calisher CH, et al. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Mol Biol Rev. 2008; 72: 457–470. doi: 10.1128/MMBR.00004-08 PMID: 18772285

2. Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases. Nature. 2007; 447: 279–283. doi: 10.1038/nature05775 PMID: 17507975

3. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, et al. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature. 1999; 397: 436–441. doi: 10.1038/17130 PMID: 9989410

4. Lemey P, Pybus OG, Wang B, Saksena NK, Salemi M, Vandamme A-M. Tracing the origin and history of the HIV-2 epidemic. Proc Natl Acad Sci USA. 2003; 100: 6588–6592. doi: 10.1073/pnas.0936469100 PMID: 12743376

5. Coffin JM, Hughes SH, Varmus HE. Retroviruses. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1997.

6. Troyer JL, VandeWoude S, Pecon-Slattery J, McIntosh C, Franklin S, Antunes A, et al. FIV cross-species transmission: an evolutionary prospective. Vet Immunol Immunopathol. 2008; 123: 159–166. doi: 10.1016/j.vetimm.2008.01.023 PMID: 18299153

7. Locatelli S, Peeters M. Cross-species transmission of simian retroviruses: how and why they could lead to the emergence of new diseases in the human population. AIDS. 2012; 26: 659–673. doi: 10.1097/QAD.0b013e328350fb68 PMID: 22441170

8. Minardi da Cruz JC, Singh DK, Lamara A, Chebloune Y. Small ruminant lentiviruses (SRLVs) break the species barrier to acquire new host range. Viruses. 2013; 5: 1867–1884. doi: 10.3390/v5071867 PMID: 23881276

9. Denner J. Transspecies transmissions of retroviruses: new cases. Virology. 2007; 369: 229–233. doi: 10.1016/j.virol.2007.07.026 PMID: 17870141

10. Niewiadomska AM, Gifford RJ. The extraordinary evolutionary history of the reticuloendotheliosis viruses. Plos Biol. 2013; 11: e1001642. doi: 10.1371/journal.pbio.1001642 PMID: 24013706

11. Gifford R, Tristem M. The evolution, distribution and diversity of endogenous retroviruses. Virus Genes. 2003; 26: 291–315. PMID: 12876457

12. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001; 409: 860–921. doi: 10.1038/35057062 PMID: 11237011

13. Mayer J, Blomberg J, Seal RL. A revised nomenclature for transcribed human endogenous retroviral loci. Mobile DNA. 2011; 2: 7. doi: 10.1186/1759-8753-2-7 PMID: 21542922

14. Magiorkinis G, Blanco-Melo D, Belshaw R. The decline of human endogenous retroviruses: extinction and survival. Retrovirology. 2015; 12: 8. doi: 10.1186/s12977-015-0136-x PMID: 25640971

15. Katzourakis A, Rambaut A, Pybus OG. The evolutionary dynamics of endogenous retroviruses. Trends in Microbiology. 2005; 13: 463–468. doi: 10.1016/j.tim.2005.08.004 PMID: 16109487

16. Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet. 2012; 13: 283–296. doi: 10.1038/nrg3199 PMID: 22421730

17. Gifford RJ. Viral evolution in deep time: lentiviruses and mammals. Trends Genet. 2012; 28: 89–100. doi: 10.1016/j.tig.2011.11.003 PMID: 22197521

18. Henzy JE, Gifford RJ, Johnson WE, Coffin JM. A novel recombinant retrovirus in the genomes of modern birds combines features of avian and mammalian retroviruses. Journal of Virology. 2014; 88: 2398–2405. doi: 10.1128/JVI.02863-13 PMID: 24352464

19. Hayward A, Grabherr M, Jern P. Broad-scale phylogenomics provides insights into retrovirus-host evolution. Proc Natl Acad Sci USA. 2013. doi: 10.1073/pnas.1315419110

20. Hayward A, Cornwallis CK, Jern P. Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. Proceedings of the National Academy of Sciences. 2015; 112: 464–469. doi: 10.1073/pnas.1414980112

21. Weiss RA. The discovery of endogenous retroviruses. Retrovirology. 2006; 3: 67. doi: 10.1186/1742-4690-3-67 PMID: 17018135

22. Mang R, Maas J, van der Kuyl AC, Goudsmit J. Papio cynocephalus endogenous retrovirus among old world monkeys: evidence for coevolution and ancient cross-species transmissions. Journal of Virology. 2000; 74: 1578–1586. PMID: 10627573

23. van der Kuyl AC, Dekker JT, Goudsmit J. Distribution of baboon endogenous virus among species of African monkeys suggests multiple ancient cross-species transmissions in shared habitats. Journal of Virology. 1995; 69: 7877–7887. PMID: 7494300

24. Yohn CT, Jiang Z, McGrath SD, Hayden KE, Khaitovich P, Johnson ME, et al. Lineage-specific expansions of retroviral insertions within the genomes of African great apes but not humans and orangutans. Plos Biol. 2005; 3: e110. doi: 10.1371/journal.pbio.0030110 PMID: 15737067

25. Polavarapu N, Bowen NJ, McDonald JF. Identification, characterization and comparative genomics of chimpanzee endogenous retroviruses. Genome Biol. 2006; 7: R51. doi: 10.1186/gb-2006-7-6-r51 PMID: 16805923

26. Gilbert C, Maxfield DG, Goodman SM, Feschotte C. Parallel germline infiltration of a lentivirus in two Malagasy lemurs. PLoS Genet. 2009; 5: e1000425. doi: 10.1371/journal.pgen.1000425 PMID: 19300488

27. Wang Y, Liška F, Gosele C, Šedová L, Křen V, Křenová D, et al. A novel active endogenous retrovirus family contributes to genome variability in rat inbred strains. Genome Research. 2010; 20: 19–27. doi: 10.1101/gr.100073.109 PMID: 19887576

28. Escalera-Zamudio M, Mendoza MLZ, Heeger F, Loza-Rubio E, Rojas-Anaya E, Méndez-Ojeda ML, et al. A novel endogenous betaretrovirus in the common vampire bat (Desmodus rotundus) suggests multiple independent infection and cross-species transmission events. Journal of Virology. 2015; 89: 5180–5184. doi: 10.1128/JVI.03452-14 PMID: 25717107

29. MARTIN J, Herniou E, Cook J, O'Neill RW, Tristem M. Interclass transmission and phyletic host tracking in murine leukemia virus-related retroviruses. Journal of Virology. 1999; 73: 2442–2449. PMID: 9971829

30. van der Kuyl AC, Dekker JT, Goudsmit J. Discovery of a new endogenous type C retrovirus (FcEV) in cats: evidence for RD-114 being an FcEV(Gag-Pol)/baboon endogenous virus BaEV(Env) recombinant. Journal of Virology. 1999; 73: 7994–8002. PMID: 10482547

31. Henzy JE, Johnson WE. Pushing the endogenous envelope. Philosophical Transactions of the Royal Society B: Biological Sciences. 2013; 368: 20120506. doi: 10.1098/rstb.2012.0506

32. Tarlinton R, Meers J, Young P. Biology and evolution of the endogenous koala retrovirus. Cell Mol Life Sci. 2008; 65: 3413–3421. doi: 10.1007/s00018-008-8499-y PMID: 18818870

33. Magiorkinis G, Gifford RJ, Katzourakis A, De Ranter J, Belshaw R. Env-less endogenous retroviruses are genomic superspreaders. Proc Natl Acad Sci USA. National Acad Sciences; 2012; 109: 7385–7390. doi: 10.1073/pnas.1200913109 PMID: 22529376

34. Mata H, Gongora J, Eizirik E, Alves BM, Soares MA, Ravazzolo AP. Identification and characterization of diverse groups of endogenous retroviruses in felids. Retrovirology. 2015; 12: 26. doi: 10.1186/s12977-015-0152-x PMID: 25808580

35. Calisher CH, Childs JE, Field HE, Holmes KV, Schountz T. Bats: Important Reservoir Hosts of Emerging Viruses. Clinical Microbiology Reviews. 2006; 19: 531–545. doi: 10.1128/CMR.00017-06 PMID: 16847084

36. Wong S, Lau S, Woo P, Yuen K-Y. Bats as a continuing source of emerging infections in humans. Rev Med Virol. 2007; 17: 67–91. doi: 10.1002/rmv.520 PMID: 17042030

37. Omatsu T, Watanabe S, Akashi H, Yoshikawa Y. Biological characters of bats in relation to natural reservoir of emerging viruses. Comparative Immunology, Microbiology and Infectious Diseases. 2007; 30: 357–374. doi: 10.1016/j.cimid.2007.05.006 PMID: 17706776

38. Luis AD, Hayman DTS, O'Shea TJ, Cryan PM, Gilbert AT, Pulliam JRC, et al. A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? Proceedings of the Royal Society B: Biological Sciences. 2013; 280: 20122753. doi: 10.1098/rspb.2012.2753 PMID: 23378666

39. Brook CE, Dobson AP. Bats as "special" reservoirs for emerging zoonotic pathogens. Trends in Microbiology. 2015; 23: 172–180. doi: 10.1016/j.tim.2014.12.004 PMID: 25572882

40. Moratelli R, Calisher CH. Bats and zoonotic viruses: can we confidently link bats with emerging deadly viruses? Mem Inst Oswaldo Cruz. 2015. doi: 10.1590/0074-02760150048

41. Smith I, Wang L-F. Bats and their virome: an important source of emerging viruses capable of infecting humans. Curr Opin Virol. 2013; 3: 84–91. doi: 10.1016/j.coviro.2012.11.006 PMID: 23265969

42. Zhuo X, Rho M, Feschotte C. Genome-wide characterization of endogenous retroviruses in the bat Myotis lucifugus reveals recent and diverse infections. Journal of Virology. 2013; 87: 8493–8501. doi: 10.1128/JVI.00892-13 PMID: 23720713

43. Pontius JU, Mullikin JC, Smith DR, Agencourt Sequencing Team, Lindblad-Toh K, Gnerre S, et al. Initial sequence and comparative analysis of the cat genome. Genome Research. 2007; 17: 1675–1689. doi: 10.1101/gr.6380007 PMID: 17975172

44. Cho YS, Hu L, Hou H, Lee H, Xu J, Kwon S, et al. The tiger genome and comparative analysis with lion and snow leopard genomes. Nat Commun. 2013; 4: 2433. doi: 10.1038/ncomms3433 PMID: 24045858

45. Seim I, Fang X, Xiong Z, Lobanov AV, Huang Z, Ma S, et al. Genome analysis reveals insights into physiology and longevity of the Brandt's bat Myotis brandtii. Nat Commun. 2013; 4: 2212. doi: 10.1038/ncomms3212 PMID: 23962925

46. Zhang G, Cowled C, Shi Z, Huang Z, Bishop-Lilly KA, Fang X, et al. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. Science. 2013; 339: 456–460. doi: 10.1126/science.1230835 PMID: 23258410

47. Yuhki N, Mullikin JC, Beck T, Stephens R, O'Brien SJ. Sequences, Annotation and Single Nucleotide Polymorphism of the Major Histocompatibility Complex in the Domestic Cat. Ellegren H, editor. PLoS ONE. 2008; 3: e2674. doi: 10.1371/journal.pone.0002674.t006 PMID: 18629345

48. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA. 2015; 6: 11. doi: 10.1186/s13100-015-0041-9 PMID: 26045719

49. Song N, Jo H, Choi M, Kim J-H, Seo HG, Cha S-Y, et al. Identification and classification of feline endogenous retroviruses in the cat genome using degenerate PCR and in silico data analysis. J Gen Virol. 2013. doi: 10.1099/vir.0.051862–0

50. Martoglio B, Graf R, Dobberstein B. Signal peptide fragments of preprolactin and HIV-1 p-gp160 interact with calmodulin. The EMBO Journal. 1997; 16: 6636–6645. doi: 10.1093/emboj/16.22.6636 PMID: 9362478

51. Corbet S, Muller-Trutwin MC, Versmisse P, Delarue S, Ayouba A, Lewis J, et al. env sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area. Journal of Virology. 2000; 74: 529–534. doi: 10.1128/JVI.74.1.529–534.2000 PMID: 10590144

52. Huet T, Cheynier R, Meyerhans A, Roelants G, Wain-Hobson S. Genetic organization of a chimpanzee lentivirus related to HIV-1. Nature. 1990; 345: 356–359. doi: 10.1038/345356a0 PMID: 2188136

53. Pancino G, Ellerbrok H, Sitbon M, Sonigo P. Conserved framework of envelope glycoproteins among lentiviruses. Curr Top Microbiol Immunol. 1994; 188: 77–105. PMID: 7924431

54. Meredith RW, Janecka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, et al. Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. Science. 2011; 334: 521–524. doi: 10.1126/science.1211028 PMID: 21940861

55. Hedges SB, Dudley J, Kumar S. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics. 2006; 22: 2971–2972. doi: 10.1093/bioinformatics/btl505 PMID: 17021158

56. Shedlock AM, Okada N. SINE insertions: powerful tools for molecular systematics. Bioessays. 2000; 22: 148–160. doi: 10.1002/(SICI)1521-1878(200002)22:2<148::AID-BIES6>3.0.CO;2-Z PMID: 10655034

57. Ray DA, Xing J, Salem A-H, Batzer MA. SINEs of a nearly perfect character. Syst Biol. 2006; 55: 928–935. doi: 10.1080/10635150600865419 PMID: 17345674

58. Maeda N, Kim HS. Three independent insertions of retrovirus-like sequences in the haptoglobin gene cluster of primates. Genomics. 1990; 8: 671–683. PMID: 2177446

59. Johnson WE, Coffin JM. Constructing primate phylogenies from ancient retrovirus sequences. Proc Natl Acad Sci USA. 1999; 96: 10254–10260. PMID: 10468595

60. Bashir A, Ye C, Price AL, Bafna V. Orthologous repeats and mammalian phylogenetic inference. Genome Research. 2005; 15: 998–1006. doi: 10.1101/gr.3493405 PMID: 15998912

61. Johnson WE. The Late Miocene Radiation of Modern Felidae: A Genetic Assessment. Science. 2006; 311: 73–77. doi: 10.1126/science.1122277 PMID: 16400146

62. Stadelmann B, Lin LK, Kunz TH, Ruedi M. Molecular phylogeny of New World Myotis (Chiroptera, Vespertilionidae) inferred from mitochondrial and nuclear DNA genes. Molecular Phylogenetics and Evolution. 2007; 43: 32–48. doi: 10.1016/j.ympev.2006.06.019 PMID: 17049280

63. Miller-Butterworth CM, Murphy WJ, O'Brien SJ, Jacobs DS, Springer MS, Teeling EC. A family matter: conclusive resolution of the taxonomic position of the long-fingered bats, miniopterus. Molecular Biology and Evolution. 2007; 24: 1553–1561. doi: 10.1093/molbev/msm076 PMID: 17449895

64. Lack JB, Roehrs ZP, Stanley CE Jr, Ruedi M, Van Den Bussche RA. Molecular phylogenetics of Myotis indicate familial-level divergence for the genus Cistugo (Chiroptera). Journal of Mammalogy. 2010; 91: 976–992. doi: 10.1644/09-MAMM-A-192.1

65. Agnarsson I, Zambrana-Torrelio CM, Flores-Saldana NP, May-Collado LJ. A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). PLoS Curr. 2011; 3: RRN1212. doi: 10.1371/currents.RRN1212 PMID: 21327164

66. Dangel AW, Baker BJ, Mendoza AR, Yu CY. Complement component C4 gene intron 9 as a phylogenetic marker for primates: long terminal repeats of the endogenous retrovirus ERV-K(C4) are a molecular clock of evolution. Immunogenetics. 1995; 42: 41–52. PMID: 7797267

67. Vitte C, Panaud O, Quesneville H. LTR retrotransposons in rice (Oryza sativa, L.): recent burst amplifications followed by rapid DNA loss. BMC Genomics. BioMed Central Ltd; 2007; 8: 218. doi: 10.1186/1471-2164-8-218 PMID: 17617907

68. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 1997; 13: 335–340. doi: 10.1016/S0168-9525(97)01181-5 PMID: 9260521

69. Pace JK, Gilbert C, Clark MS, Feschotte C. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proc Natl Acad Sci USA. 2008; 105: 17023–17028. doi: 10.1073/pnas.0806548105 PMID: 18936483

70. Kumar S, Subramanian S. Mutation rates in mammalian genomes. Proc Natl Acad Sci USA. 2002; 99: 803–808. doi: 10.1073/pnas.022629899 PMID: 11792858

71. Hughes JF. Human Endogenous Retroviral Elements as Indicators of Ectopic Recombination Events in the Primate Genome. Genetics. 2005; 171: 1183–1194. doi: 10.1534/genetics.105.043976 PMID: 16157677

72. Kijima TE, Innan H. On the Estimation of the Insertion Time of LTR Retrotransposable Elements. Molecular Biology and Evolution. 2010; 27: 896–904. doi: 10.1093/molbev/msp295 PMID: 19955475

73. Slattery JP, Franchini G, Gessain A. Genomic evolution, patterns of global dissemination, and inter-species transmission of human and simian T-cell leukemia/lymphotropic viruses. Genome Research. 1999; 9: 525–540. PMID: 10400920

74. Fernández-Medina RD, Ribeiro JMC, Carareto CMA, Velasque L, Struchiner CJ. Losing identity: structural diversity of transposable elements belonging to different classes in the genome of Anopheles gambiae. BMC Genomics. 2012; 13: 272. doi: 10.1186/1471-2164-13-272 PMID: 22726298

75. Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, Burt A, et al. Long-term reinfection of the human genome by endogenous retroviruses. Proc Natl Acad Sci USA. 2004; 101: 4894–4899. doi: 10.1073/pnas.0307800101 PMID: 15044706

76. Belshaw R, Katzourakis A, Paces J, Burt A, Tristem M. High copy number in human endogenous retrovirus families is associated with copying mechanisms in addition to reinfection. Molecular Biology and Evolution. 2005; 22: 814–817. doi: 10.1093/molbev/msi088 PMID: 15659556

77. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution. 2007; 24: 1586–1591. doi: 10.1093/molbev/msm088 PMID: 17483113

78. Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Molecular Biology and Evolution. 1998; 15: 568–573. PMID: 9580986

79. Ribet D, Harper F, Dupressoir A, Dewannieux M, Pierron G, Heidmann T. An infectious progenitor for the murine IAP retrotransposon: Emergence of an intracellular genetic parasite from an ancient retrovirus. Genome Research. 2008; 18: 597–609. doi: 10.1101/gr.073486.107 PMID: 18256233

80. Meyerson NR, Sawyer SL. Two-stepping through time: mammals and viruses. Trends in Microbiology. 2011; 19: 286–294. doi: 10.1016/j.tim.2011.03.006 PMID: 21531564

81. Daugherty MD, Malik HS. Rules of engagement: molecular insights from host-virus arms races. Annu Rev Genet. 2012; 46: 677–700. doi: 10.1146/annurev-genet-110711-155522 PMID: 23145935

82. Demogines A, Abraham J, Choe H, Farzan M, Sawyer SL. Dual host-virus arms races shape an essential housekeeping protein. Plos Biol. 2013; 11: e1001571. doi: 10.1371/journal.pbio.1001571 PMID: 23723737

83. Gaudin TJ, Emry RJ, Wible JR. The phylogeny of living and extinct pangolins (Mammalia, Pholidota) and associated taxa: a morphology based analysis. Journal of mammalian evolution. 2009. doi: 10.1007/s10914-009-9119-9

84. Ancillotto L, Serangeli MT, Russo D. Curiosity killed the bat: domestic cats as bat predators. Mammalian Biology-Zeitschrift für . . .. 2013. doi: 10.1016/j.mambio.2013.01.003

85. Phillips S, Coburn D, James R. An observation of cat predation upon an eastern blossom bat Syconycteris australis. Australian Mammalogy. 2001.

86. Scrimgeour J, Beath A, Swanney M. Cat predation of short-tailed bats (Mystacina tuberculata rhyocobia) in Rangataua Forest, Mount Ruapehu, Central North Island, New Zealand. New Zealand Journal of Zoology. 2012; 39: 257–260. doi: 10.1080/03014223.2011.649770

87. Woods M, McDonald RA, Harris S. Predation of wildlife by domestic cats Felis catus in Great Britain. Mammal review. 2003.

88. Grzimek B, Kleiman DG, Schlager N, Geist V, Olendorf D, McDade MC, et al. Grzimek's Animal Life Encyclopedia: Mammals I-V. Gale / Cengage Learning; 2003.

89. Dacheux L, Larrous F, Mailles A. European bat lyssavirus transmission among cats, Europe. Emerging infectious . . .. 2009. doi: 10.3201/eid1502.080637

90. Dewannieux M, Dupressoir A, Harper F, Pierron G, Heidmann T. Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. Nat Genet. 2004; 36: 534–539. doi: 10.1038/ng1353 PMID: 15107856

91. Mietz JA, Grossman Z, Lueders KK, Kuff EL. Nucleotide sequence of a complete mouse intracisternal A-particle genome: relationship to known aspects of particle assembly and function. Journal of Virology. 1987; 61: 3020–3029. PMID: 3041022

92. David VA, Menotti-Raymond M, Wallace AC, Roelke M, Kehler J, Leighty R, et al. Endogenous Retrovirus Insertion in the KIT Oncogene Determines White and White spotting in Domestic Cats. G3 (Bethesda). 2014. doi: 10.1534/g3.114.013425

93. Ndung'u T, Weiss RA. On HIV diversity. AIDS. 2012; 26: 1255–1260. doi: 10.1097/QAD.0b013e32835461b5 PMID: 22706010

94. Moosman PR Jr, Thomas HH. Diet of the widespread insectivorous bats Eptesicus fuscus and Myotis lucifugus relative to climate and richness of bat communities. Journal of . . .. 2012. doi: 10.1644/11-MAMM-A-274.1

95. Hutson AM, Mickleburgh SP, Racey PA. Microchiropteran Bats. IUCN; 2001.

96. Katzourakis A, Magiorkinis G, Lim AG, Gupta S, Belshaw R, Gifford R. Larger Mammalian Body Size Leads to Lower Retroviral Activity. PLoS Pathog. Public Library of Science; 2014; 10: e1004214. doi: 10.1371/journal.ppat.1004214 PMID: 25033295

97. Hayman DTS, Bowen RA, Cryan PM, McCracken GF, O'Shea TJ, Peel AJ, et al. Ecology of zoonotic infectious diseases in bats: current knowledge and future directions. Zoonoses Public Health. 2013; 60: 2–21. doi: 10.1111/zph.12000 PMID: 22958281

98. O'Shea TJ, Cryan PM, Cunningham AA, Fooks AR, Hayman DTS, Luis AD, et al. Bat flight and zoonotic viruses. Emerging Infect Dis. 2014; 20: 741–745. doi: 10.3201/eid2005.130539 PMID: 24750692

99. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA. Database indexing for production MegaBLAST searches. Bioinformatics. 2008; 24: 1757–1764. doi: 10.1093/bioinformatics/btn322 PMID: 18567917

100. Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics. 2008; 9: 18. doi: 10.1186/1471-2105-9-18 PMID: 18194517

101. Steinbiss S, Willhoeft U, Gremme G, Kurtz S. Fine-grained annotation and classification of de novo predicted LTR retrotransposons. Nucleic Acids Research. 2009; 37: 7002–7013. doi: 10.1093/nar/gkp759 PMID: 19786494

102. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26: 841–842. doi: 10.1093/bioinformatics/btq033 PMID: 20110278

103. Smit A, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. [Internet].

104. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010; 59: 307–321. doi: 10.1093/sysbio/syq010 PMID: 20525638

105. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004; 32: 1792–1797. doi: 10.1093/nar/gkh340 PMID: 15034147

106. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Molecular Biology and Evolution. 2010; 27: 221–224. doi: 10.1093/molbev/msp259 PMID: 19854763

107. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980; 16: 111–120. PMID: 7463489

108. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003; 52: 696–704. PMID: 14530136

109. Löytynoja A, Goldman N. An algorithm for progressive multiple alignment of sequences with insertions. Proc Natl Acad Sci USA. 2005; 102: 10557–10562. doi: 10.1073/pnas.0409137102 PMID: 16000407

110. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 2012; 9: 772. doi: 10.1038/nmeth.2109 PMID: 22847109

111. Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol. 2012; 61: 1061–1067. doi: 10.1093/sysbio/sys062 PMID: 22780991

112. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011; 7: e1002195. doi: 10.1371/journal.pcbi.1002195 PMID: 22039361

113. Garwood F. Fiducial limits for the Poisson distribution. Biometrika. 1936. doi: 10.2307/2333958