# Through the Genetic Looking Glass: Paleoviruses and the Present

**By Skylar Walters**

**Abstract**

Paleovirology is an emerging field combining genetics, virology, and evolutionary biology to understand the modern-day impacts of ancient viruses. Most commonly, we study ancient viruses through endogenous viral elements, or EVEs, which are full or fragmented viral genetic sequences that have become part of a host genome through viral insertion into host germline cells. Analyzing the presence and absence of these EVEs in related species provides us a fuller understanding of viral evolution and can identify why certain viruses have specific qualities, such as virulence, in different host species. When these EVEs enter the genome, they typically lose their coding potential. However, some retain coding abilities and are co-opted for productive uses in the body, notably syncytin, a protein that is essential for mammal placental development. Such coding EVEs have also been linked to cancer risk factors, thus the complex relationships between paleoviruses in the genome can be harnessed for both productive and destructive uses in the body.

This paper provides a multifaceted view of viral fossils and EVEs: one that relates to the power of viruses to persist for millennia and one that relates their importance to humans as a species– and life as a whole.

## Introduction

The field of paleovirology describes the phenomena of investigating the modern impacts of ancient viruses. Due to their often-extinct nature, direct observation of paleoviruses is difficult. Thus far, direct paleovirus sampling has been limited to finding samples frozen in permafrost, such as the recently-sequenced 48,500 year old *pandoravirus* isolated from Siberian permafrost in 2022 (Alempic et al., 2023). The more common route of studying paleoviruses revolves around indirect observation. Endogenous viral elements (EVEs) are DNA sequences originating from viruses that are housed within the germline cells of a host. These EVEs can date back millions of years in the genetic record, acting as a method of observation for ancient viruses.

Modern paleovirus research primarily revolves around two outlets: tracing modern-day virus lineages through the genetic record and understanding the role of paleoviruses in bodily function. These fields are essential to modern virus research in that they inform our understanding of the treatment, prevention, and virulence of illnesses. Additionally, paleovirus remnants in the body pose a means of establishing function and treatment of disease. This paper thus aims to provide an introduction to the far past through viral history while relating these findings to modern biological sciences.

**Section 1: EVE Formation Mechanisms**

When searching for the oldest viruses, our methods revolve around locating and dating EVEs. In mammal genomes, endogenous retroviral elements, or REVs, are common due to the nature of retroviruses integrating into the host genome to complete their life cycles. When a retrovirus infects a cell, it uses its reverse transcriptase enzyme to create a DNA transcript of its RNA, known as complementary DNA (cDNA). The virus then uses its integrase enzyme to remove 2 nucleotides from each 3' end of the cDNA, while two hydroxyl groups are added to the location on the host cell's DNA where the cDNA will join. Then, gap repair enzymes from the host cell attach the cDNA to the host cell DNA, adding the viral genetic sequence to the host's genetic sequence (Skalka & Katz, 2005). If this process occurs in a germline cell, the integrated sequence can be passed onto offspring, conserving the viral sequence in their genome. As this continues over generations, the viral genome can become a permanent part of the species' genome. Through this process, retroviruses can be encapsulated within a host species genome for millions of years, creating a paleovirus "fossil" (Katzourakis & Gifford, 2010).

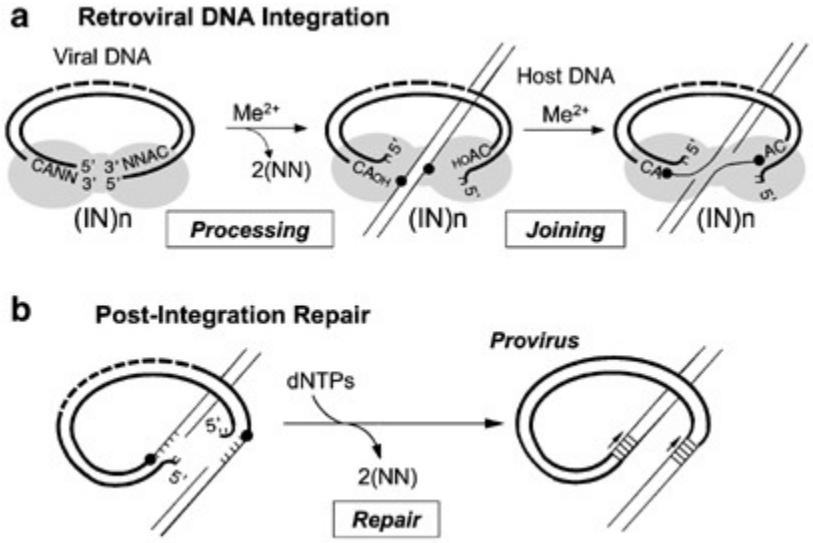


*Fig. 1: Retroviral integration into a host-cell genome (Skalka & Katz, 2005)*

In non-retroviruses, this process is more difficult, as they do not code for reverse transcriptases or integrases. Therefore, the formation of a non-retroviral EVE necessitates either a) the presence of endogenous enzymes in cells that can reverse transcribe the viral RNA or b) simultaneous infection by a retrovirus whose reverse transcriptase can create the viral DNA. After this viral DNA is created, it can be inserted into the host cell in the same manner as retroviral DNA, creating a non-retroviral EVE. The extent of the non-retroviral entry mechanism

is not fully understood, however, and could be virus-dependent (Houe, Bonizzoni, and Failloux, 2019).

## Section 2: Case Studies of Viral Evolution and Modern Implications

Studying the evolutionary histories of modern viruses from paleoviruses allows us to understand the progression of diseases through history, in turn allowing for greater knowledge on host-pathogen interactions and disease outbreak mitigation.

### a. Bornaviruses

The first non-retroviral fossil was discovered in 2010 by a research group in Japan studying bornaviruses (Horie et al., 2011). Bornaviruses are a family of viruses that cause severe neurological illness characterized by a loss of muscle control and severe depression. Though they are uncommon in humans, bornaviruses are a unique subject of experimentation since they are the only animal RNA viruses capable of surviving in the cell nucleus.

To understand why bornaviruses are capable of nucleus survival, the group searched for bornavirus proteins that were similar to the proteins in the human nucleus. They found a high degree of similarity and alignment between two genes in humans, LOC340900 and LOC55096, and the N-gene in the bornaviruses, suggesting that the human genes were endogenized fragments of ancient bornavirus N-genes. These two genes were dubbed EBLNs, or Borna-like N elements.

The researchers then conducted a BLAST search for additional EBLNs, revealing two more EBLNs in the human genome and in many other mammals. Interestingly, all of the primate EBLNs except for one coded for mRNAs, suggesting that the viral proteins play a role in coding in the modern genome. Researchers then expanded this search for EBLNs for reptilian and avian bornaviruses, constructing a phylogenetic tree (Fig. 2).

The primate EBLNs, in green, and the murine EBLNs, in dark red, showed high similarity, indicating that their bornavirus sequences infected a common ancestor and spread through millions of years of evolution, while the species in blue and bright red gained their sequences through more recent infection. By cross referencing which species carry the EVE, researchers can place the emergence of the virus to a specific point in time. For example, if species A and B diverged 25,000 years ago, and species A has the gene but species B does not, we can determine that the gene is less than 25,000 years old. This technique and the researchers' phylogenetic tree dated at upwards of 40 million years old.

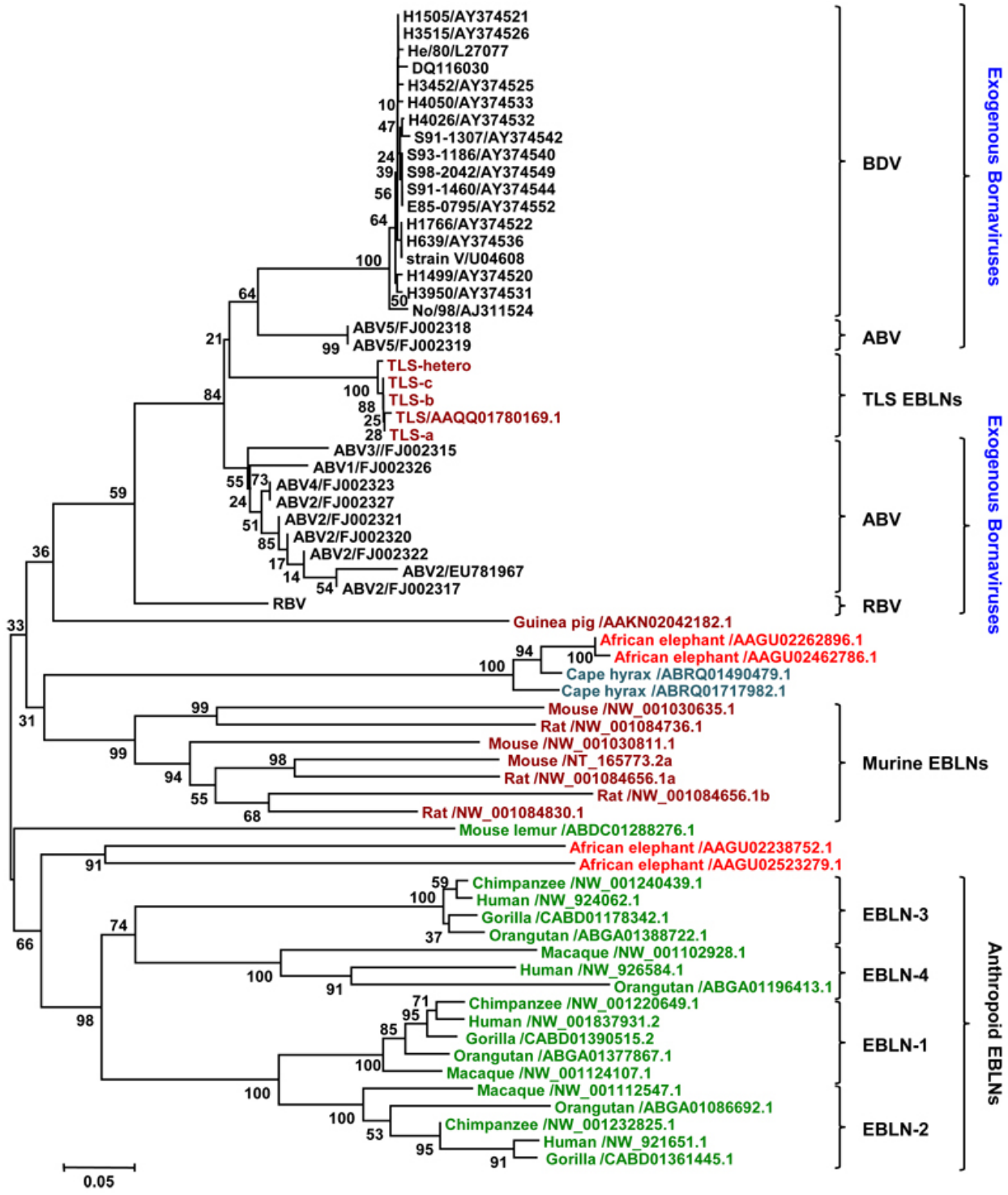


*Fig. 2: Phylogenetic tree of organisms and their EBLN sequences (Horie et al., 2011)*

### b. Human Immunodeficiency Virus (HIV)

HIV is widely accepted to have emerged from the primate virus Simian Immunodeficiency Virus (SIV), which likely began infecting humans during the early 1900s (Korber et al., 2000). Therefore, studying the evolution of HIV in humans is aided strongly by studying the evolution of SIV. Within this, scientists hope to answer the key question of why HIV is so virulent in humans, yet SIV is almost completely benign in its natural primate hosts.

Compared to the previously described paleoviruses, SIV is more recent in the genetic record. Using SIV EVE presence/absence analysis in various primates and molecular clocks, a 2009 study placed the time of SIV emergence in the genetic record at the late 1400s, at which point it began infecting chimpanzees (Wertheim & Worobey, 2009).

Prior to this study, the virulence/avirulence contrast between HIV and SIV was typically explained by SIV having coevolved with non-human primates for millions of years, losing its virulence over time. However, the suggestion of SIV being a young virus disproves this leading theory, creating new outlets for research into what makes HIV malignant.

Although this paleovirus created many more questions in HIV virulence, it proves a critical point about the importance of paleovirus research: understanding the evolution of viruses alongside animals can provide us with critical answers to questions regarding infectivity, treatment, and prevention.

## Section 3: Applications of Paleoviruses in the Human Genome

Upwards of 8% of the human genome consists of endogenous retroviral elements (REVs), and when we include REV fragments and derivative sequences, this number rises to almost half of our genome (Bannert & Kurth, 2004; Medstrand, van de Lagemaat, & Mager, 2002). These elements are typically classified by the presence of *env*, *gag*, and *pol* genes, which originate in viruses. Thus, their presence in a non-viral organism indicates that the sequence was inserted at some time in the organism's genetic history. The vast majority of these inserted genes are non-coding. However, some sequences have maintained their coding abilities, and have been harnessed for both productive and destructive use in the body.

### a. Role in Carcinogenesis Suppression and Promotion

EVEs are a promising area for cancer research, as they have been suggested to be capable of simultaneously suppressing and promoting cancer. When comparing activation rates between healthy and cancerous tissue, researchers have found evidence that human endogenous retroviruses (HERVs) are activated more frequently in a number of cancerous tissues, including,

but not limited to, prostate, lung, bladder, ovarian, and blood cancers, seminomas, soft tissue sarcomas, and Kaposi's sarcoma (Gao, Yu, & Chen, 2021).

The exact mechanism explaining this association between HERV expression and cancers is still debated, and is likely dependent on the exact cancers and HERVs in play. Many studies have been looking into the role of the HERV *env* protein as an impetus for loss of cell regulation. A 2013 study focused attention on immunizing rats against a form of *env* protein produced by HERV-K HML-2, or HK2, the most active subtype of HERV, and found that there was a strong association between *env* suppression and reduction in tumor growth, suggesting that the *env* protein plays a significant role in the formation of tumors in renal carcinomas (Kraus et al., 2013).

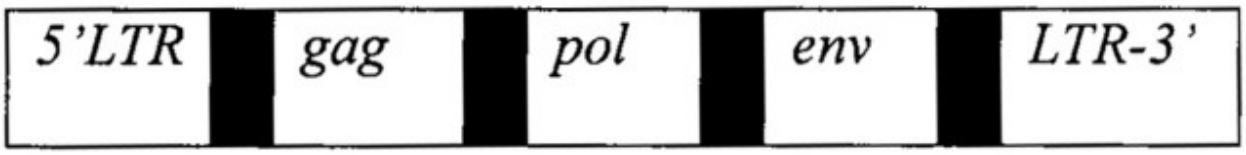


*Fig. 3: HERV structure consists of long terminal repeats (LTRs) bookmarking* gag, pol*, and* env *genes (Ryan, 2004)*

HK2 has also been proposed as a key HERV in breast cancers. Antibodies against HK2 *env* proteins are detected in early-stage breast cancer patients in higher proportions than the general population (Tavakolian, Goudarzi, & Faghihloo, 2019). This, coupled with the fact that HK2 env antibodies naturally decrease with age, suggests a link between this protein and breast cancers. As a result, HK2-env-targeting vaccines prove to be a promising outlet for future research in cancer prevention (Mastrangelo, 2018).

However, the opposite role of HERVs as a means of cancer prevention has also gained scientific attention. Products of transcribed HERVs have been found to activate innate, cellular, and humoral immunity in various cancers, including breast and renal cancers (Gao, Yu, & Chen, 2021). Insertion of these HERV products in tumors have been shown to reduce tumorigenesis, suggesting that the HERVs could potentially be harnessed as a means of preventing and treating cancers (Tokuyama 2018). The mechanism behind this process, however, is poorly understood and is a prime outlet for future research.

**b. Syncytins**

EVEs have also evolved within the genome to have unambiguous function in the body, a prime example being syncytins. Syncytins are endogenous retroviral *env* genes that are estimated to be 30-45 million years old in humans (Lavialle et al., 2013). Syncytins were co-opted from

envelope proteins of a now-extinct viral lineage, with the genome preserving the envelope protein's fusing properties for purposes in placental development.

Syncytin genes produce proteins that fuse to placental cell receptors. The placenta consists of multiple cellular layers, namely the cytotrophoblast. This is the inner layer of the tissue surrounding a mammal blastula that provides an embryo with nourishment from the parent. As the embryo grows, this layer develops into the placenta to continue providing the growing fetus with nutrients. Syncytin proteins fuse these layers of cytotrophoblasts, allowing the blastula to develop and provide the embryo with nutrients as it grows. This syncytial layer is also essential to selective permeability of the developing embryo. Parent immune systems are composed of white blood cells that can enter tissue by traveling through small gaps in cells. If these cells could permeate the placenta into the areas of fetal development, the foreign cell and proteins would trigger an immune response that could damage the growing blastula. However, the heavily-fused syncytium prevents this permeation, ensuring the growth of the fetus. Thus, the viral mechanism that was initially used to grow a viral envelope protein has evolved with mammals to aid in placental development (Blond et al., 2000).

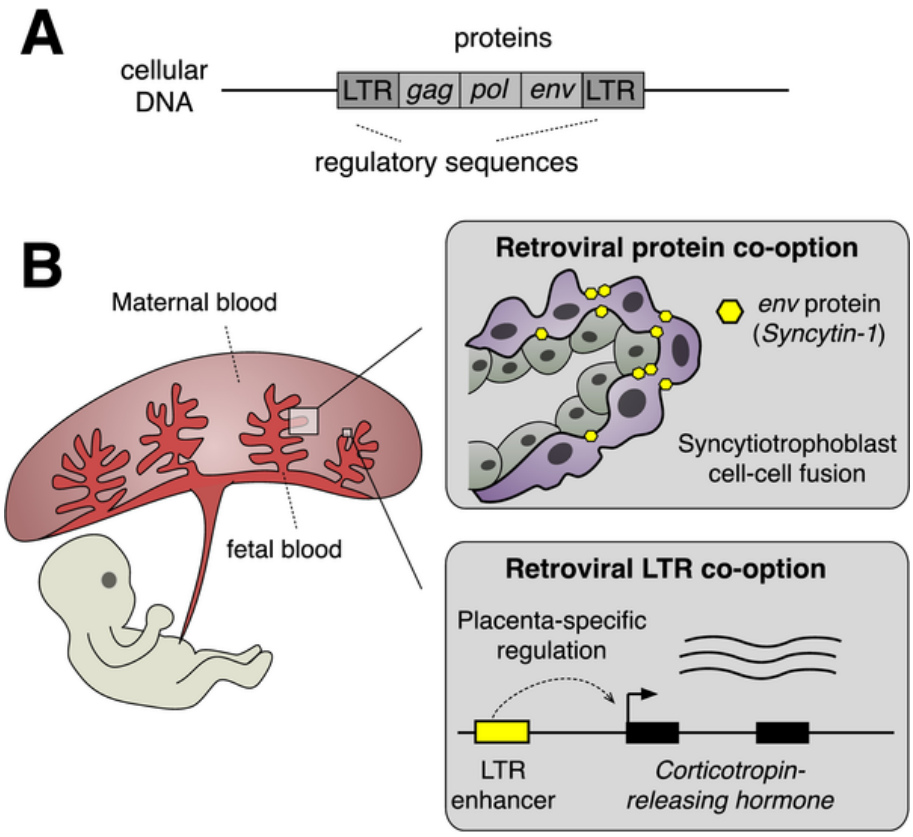


*Fig. 4: Syncytin in placental development: the syncytin env proteins fuse and close gaps between the syncytiotrophoblasts (Chuong, 2018)*

Interestingly, syncytins have independently entered mammalian genomes a number of times throughout evolutionary history, as opposed to once in an ancient ancestor to all mammals. This suggests that the development of placentation throughout mammal evolution could be closely tied to the infection and *env* co-option by the ancient syncytin-inserting virus (Lavialle et al., 2013.

**Conclusion**

Paleoviruses provide us with a novel lens into the far past. Their presence in non-viral genomes in the form of endogenous viral elements (EVEs) allows us to trace viral lineages through phylogenetic history while investigating critical differences between how viruses impact different organisms. As a result of being present in the genome, ancient viruses have the potential to be expressed in modern genes. This can be an impetus for good, such as in the expression of syncytins and activation of immune responses for cancer, but may also contribute to carcinogenesis.

There are also a number of limitations in the current study of paleoviruses. Since retroviruses are capable of entering the genome more easily than non-retroviruses, retroviruses are overrepresented in the research of paleoviruses. However, this does not mean that non-retroviruses have not had an impact on the human genome and evolution of species as a whole. Rather, they likely had just as significant of an impact, though their fossils are not left behind. Consequently, a key area for future research in paleoviruses involves locating more EVEs of non-retroviruses, as well as developing new ways to detect non-retroviruses that may not be in the form of EVEs.

Additionally, our understanding of the role of REVs in carcinogenesis and cancer prevention remains very open-ended. Often, there are direct contradictions between the promotion and suppression of cancer; for example, HK2 has been of particular interest in breast cancer due to its conflicting ties to causing tumors, but also eliciting immune responses (Kraus et al., 2013). Thus, future paleovirus research should focus on harnessing the tools within the human genome to discover the mechanisms contributing to and preventing cancer.

With viruses growing alongside life for billions of years, the intersections of ancient viruses with ancient organisms have had rippling ramifications that impact us to this day, a process that will continue to the end of organismal life itself. While we have viewed these interactions as inherently negative for most of human history, we are learning that this relationship has complexity and a degree of symbiosis that we can harness for the betterment of life as a whole.

## Works Cited

Alempic, J. M., Lartigue, A., Goncharov, A. E., Grosse, G., Strauss, J., Tikhonov, A. N., Fedorov, A. N., Poirot, O., Legendre, M., Santini, S., Abergel, C., & Claverie, J. M. (2023). An Update on Eukaryotic Viruses Revived from Ancient Permafrost. Viruses, 15(2), 564. https://doi.org/10.3390/v15020564

Bannert, N., & Kurth, R. (2004). Retroelements and the human genome: new perspectives on an old relation. Proceedings of the National Academy of Sciences of the United States of America, 101 Suppl 2(Suppl 2), 14572–14579. https://doi.org/10.1073/pnas.0404838101

Blond, J. L., Lavillette, D., Cheynet, V., Bouton, O., Oriol, G., Chapel-Fernandes, S., ... & Cosset, F. L. (2000). An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. Journal of virology, 74(7), 3321-3329.

Chuong, E. B. (2018). The placenta goes viral: Retroviruses control gene expression in pregnancy. PLoS biology, 16(10), e3000028.

Gao, Y., Yu, X. F., & Chen, T. (2021). Human endogenous retroviruses in cancer: Expression, regulation and function. Oncology letters, 21(2), 121. https://doi.org/10.3892/ol.2020.12382

Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P., Gojobori, T., Coffin, J. M., & Tomonaga, K. (2010). Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature, 463(7277), 84–87. https://doi.org/10.1038/nature08695

Houé, V., Bonizzoni, M., & Failloux, A. B. (2019). Endogenous non-retroviral elements in genomes of Aedes mosquitoes and vector competence. Emerging microbes & infections, 8(1), 542-555.

Katzourakis, A., & Gifford, R. J. (2010). Endogenous viral elements in animal genomes. PLoS genetics, 6(11), e1001191.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S., & Bhattacharya, T. (2000). Timing the ancestor of the HIV-1 pandemic strains. Science (New York, N.Y.), 288(5472), 1789–1796. https://doi.org/10.1126/science.288.5472.1789

Kraus, B., Fischer, K., Büchner, S. M., Wels, W. S., Löwer, R., Sliva, K., & Schnierle, B. S. (2013). Vaccination directed against the human endogenous retrovirus-K envelope protein

inhibits tumor growth in a murine model system. PloS one, 8(8), e72756. https://doi.org/10.1371/journal.pone.0072756

Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C., & Heidmann, T. (2013). Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 368(1626), 20120507. https://doi.org/10.1098/rstb.2012.0507

Mastrangelo, G., Pavanello, S., Fadda, E., Buja, A., & Fedeli, U. (2018). Yellow fever vaccine 17D administered to healthy women aged between 40 and 54 years halves breast cancer risk: an observational study. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP), 27(4), 303–309. https://doi.org/10.1097/CEJ.0000000000000333

Medstrand, P., van de Lagemaat, L. N., & Mager, D. L. (2002). Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome research, 12(10), 1483–1495. https://doi.org/10.1101/gr.388902

Ryan F. P. (2004). Human endogenous retroviruses in health and disease: a symbiotic perspective. Journal of the Royal Society of Medicine, 97(12), 560–565. https://doi.org/10.1177/014107680409701202

Skalka, A. M., & Katz, R. A. (2005). Retroviral DNA integration and the DNA damage response. Cell Death & Differentiation, 12(1), 971-978.

Tavakolian, S., Goudarzi, H., & Faghihloo, E. (2019). Evaluating the expression level of HERV-K env, np9, rec and gag in breast tissue. Infectious agents and cancer, 14, 42. https://doi.org/10.1186/s13027-019-0260-7

Tokuyama, M., Kong, Y., Song, E., Jayewickreme, T., Kang, I., & Iwasaki, A. (2018). ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. Proceedings of the National Academy of Sciences of the United States of America, 115(50), 12565–12572. https://doi.org/10.1073/pnas.1814589115

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